NOVEL APPLICATIONS FOR EMERGING MARKETS USING TELEVISION AS A UBIQUITOUS DEVICE

Pal, Arpan

Publication date:
2013

Document Version
Early version, also known as pre-print

NOVEL APPLICATIONS FOR EMERGING MARKETS USING TELEVISION AS A UBIQUITOUS DEVICE

DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
ELECTRONIC SYSTEMS
OF
AALBORG UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Arpan Pal
Center for TeleInFrastruktur Denmark
and
Tata Consultancy Services India

Supervisor:
Prof. Ramjee Prasad, Aalborg University, Denmark

The Assessment Committee:
Director Sudhir Dixit, HP Laboratories, India
Professor Mary Ann Weitnauer, Georgia Institute of Technology, USA
Professor Emeritus Jørgen Bach Andersen, Aalborg University, Denmark (Chairman)

Moderator:
Associate Professor Albena Mihovska, Aalborg University, Denmark

ISBN: 978-87-7152-005-7

Arpan Pal
Center for TeleInFrastruktur (CTIF)
Aalborg University
Niels Jernes Vej 12
9220 Aalborg
Denmark
Email: arpan.pal@tcs.com

Dedicated to my Parents, Wife and Daughter

Thesis title - NOVEL APPLICATIONS FOR EMERGING MARKETS USING TELEVISION AS A UBIQUITOUS DEVICE
Name of PhD student – Arpan Pal
Name and title of supervisor and any other supervisors – Ramjee Prasad, Professor, CTIF
List of published papers:

- Arpan Pal, M. Prashant, Avik Ghose, Chirabrata Bhaumik, "Home Infotainment Platform – A Ubiquitous Access Device for Masses", Proceedings on Ubiquitous Computing and Multimedia Applications (UCMA), Miyazaki, Japan, March 2010.
- Dhiman Chattopadhyay, Aniruddha Sinha, T. Chattopadhyay, Arpan Pal, "Adaptive Rate Control for H.264 Based Video Conferencing Over a Low Bandwidth Wired and Wireless Channel", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, May 2009.
- Arpan Pal and T. Chattopadhyay, "A Novel, Low-Complexity Video Watermarking Scheme for H.264", Texas Instruments Developers Conference, Dallas, Texas, March 2007.
- T. Chattopadhyay and Arpan Pal, "Two fold video encryption technique applicable to H.264 AVC", IEEE International Advance Computing Conference (IACC), Patiala, India, March 2009.
- T. Chattopadhyay, Aniruddha Sinha, Arpan Pal, Debabrata Pradhan, Soumali Roychowdhury, "Recognition of Channel Logos From Streamed Videos for Value Added Services in Connected TV", IEEE International Conference for Consumer Electronics (ICCE), Las Vegas, USA , January 2011.
- T. Chattopadhyay, Arpan Pal, Utpal Garain, "Mash up of Breaking News and Contextual Web Information: A Novel Service for Connected Television", Proceedings of 19th International Conference on Computer Communications and Networks (ICCCN), Zurich, Switzerland, August 2010.
- T. Chattopadhyay, Aniruddha Sinha, Arpan Pal, "TV Video Context Extraction", IEEE Trends and Developments in Converging Technology towards 2020 (TENCON 2011), Bali, INDONESIA, November 21-24, 2011.
- Arpan Pal, Chirabrata Bhaumik, Debnarayan Kar, Somnath Ghoshdastidar, Jasma Shukla, "A Novel On-Screen Keyboard for Hierarchical Navigation with Reduced Number of Key Strokes", IEEE International Conference on Systems, Man and Cybernetics (SMC), San Antonio, Texas, October 2009.
- Arpan Pal, Debatri Chatterjee, Debnarayan Kar, "Evaluation and Improvements of on-screen keyboard for Television and Set-top Box", IEEE International Symposium for Consumer Electronics (ISCE), Singapore, June 2011.
- Arpan Pal, M. Prashant, Avik Ghose, Chirabrata Bhaumik, "Home Infotainment Platform – A Ubiquitous Access Device for Masses", Book Chapter in Springer Communications in Computer and Information Science, Volume 75, 2010, Pages 11-19.DOI: 10.1007/978-3-642-13467-8.
- Arpan Pal, Ramjee Prasad, Rohit Gupta, "A low-cost Connected TV platform for Emerging Markets–Requirement Analysis through User Study", *Engineering Science and Technology: An International Journal (ESTIJ)*, ISSN: 2250-3498, Vol.2, No.6, December 2012.
- T. Chattopadhyay and Arpan Pal, "Watermarking for H.264 Video", EE Times Design, Signal Processing Design Line, November 2007.
- Arpan Pal, Aniruddha Sinha and Tanushyam Chattopadhyay, "Recognition of Characters from Streaming Videos", Book Chapter in book: Character Recognition, Edited by Minoru Mori, Sciyo Publications, ISBN: 978-953-307-105-3, September 2010.
- Arpan Pal, Tanushyam Chattopadhyay, Aniruddha Sinha and Ramjee Prasad, "The Context-aware Television using Logo Detection and Character Recognition", (Submitted) *Springer Journal of Pattern Analysis and Applications*
- Debatri Chatterjee, Aniruddha Sinha, Arpan Pal, Anupam Basu,, "An Iterative Methodology to Improve TV Onscreen Keyboard Layout Design Through Evaluation of user Study", Journal of Advances in Computing, Vol.2, No.5, October 2012), Scientific and Academic Publishing (SAP), p-ISSN:2163-2944, e-ISSN:2163-2979.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Abstract

According to Mark Weiser, who is known as the father of ubiquitous computing, the most profound technologies are those that disappear and disappearance of technology is the fundamental requirement for ubiquity. Ubiquitous computing is more about computing for the purpose of communication and for the purpose of information access anytime everywhere. As we enter the ubiquitous computing era, the main aspect of ubiquity will be how these computing devices disseminate information to end-users. As of today, there are three "main screens" in our life that can be used for such information dissemination – personal computer (including laptop and tablet), television and mobile (including smart phone). For emerging market countries like India, personal computers are not yet affordable to masses and most of the people are not savvy or skilled enough to operate a personal computer. Mobile phones, though being low-cost, pervasive and easy-to-use, suffer from the problem of having very small screen real-estate where very little useful information can be disseminated. Television however is a truly pervasive device that has penetrated the homes of the masses in these countries. If one could make the television connected to the internet-world in a low-cost manner, it has the potential of becoming the "Ubiquitous Computing Screen" for the home.

The emerging markets are characterized by some unique issues like low bandwidth / low Quality-of-Service (QoS) of the available wireless networks, extreme cost-consciousness of the users and lack of computer literacy among masses. The technological challenges arising from these issues are lack of reliable and low complexity protocols and security schemes for multimedia content delivery over low-bandwidth low quality wireless networks and the lack of low-cost and easy to use solutions for seamlessly blending Internet with broadcasting content.

To this end, in this thesis a novel application development framework is proposed first on top of a low-cost over-the-top box that uses television as a ubiquitous device with a focus on emerging markets. Then an end-to-end solution with relevant applications is created using the framework for deployment in the real-world to gather user feedback. The user feedback is analyzed in context of the above-mentioned challenges typical of developing countries and is translated into a set of problem requirements.

Based on these requirements, the thesis further introduces a set of novel low complexity protocols and security schemes that addresses the low bandwidth / low-QoS network issues and low-computing power issues of the device. The thesis also proposes an intelligent blending of broadcast television and internet through channel logo detection and optical character recognition for addressing the ease-of-use issue. The thesis finally introduces a novel text entry scheme in television using infra-red remote controls through an innovative on-screen keyboard layout to address ease-of-use issue for non-computer-savvy users.

Dansk Resume

Ifølge Mark Weiser, der er kendt som grundlæggeren af ”ubiquitous computing”, er de mest dybsindige teknologier dem, der forsvinder og forsvinden af teknologi er det grundlæggende krav for ”ubiquity”. ”Ubiquitous computing” handler om beregninger med henblik på kommunikation og med henblik på adgang til informationer når som helst, hvor som helst. Når vi træder ind i ”ubiquitous computing” æraen, vil det vigtigste aspekt af ”ubiquity” være, hvordan disse computerenheder formidler information til slutbrugerne. I dag er der tre "main screens" i vores liv, der kan bruges til en sådan formidling – PC’er (herunder laptop og tablet), tv og mobil (herunder smartphone). For ”emerging markets” lande som Indien, er PC’er endnu uoverkommelige for mange, og de fleste er ikke kyndige eller dygtige nok til at betjene en PC. Mobiltelefoner er billige, udbredte og nemme at bruge, men begrænses af at have en lille skærm, hvor kun lidt information kan formidles. Tv er imidlertid meget udbredt og kan findes i mange hjem i disse lande. Hvis man kunne få fjernsynet tilsluttet internettet på en billig måde, har det potentiale til at blive den "Ubiquitous Computing Screen" i hjemmet.

De nye markeder er præget af nogle unikke problemer som lav båndbredde / lav Quality-of-Service (QoS) af de tilgængelige trådløse netværk, ekstrem omkostningsbevidsthed af forbrugerne og manglende IT-færdigheder blandt befolkningen. De teknologiske udfordringer som følge af disse problemer, er mangel på pålidelige og lav-kompleks protokoller og sikkerhed for multimedieindhold leveret over lav båndbredde, lav kvalitets trådløse netværk og manglen på billige og nemme løsninger til problemfrit at mikse internettet med radio-/tv-indhold.

Til dette formål, er der i denne rapport foreslået rammer for applikationsudvikling, først på toppen af en billig ”over-the-top-box”, der bruger fjernsynet som en tilgængelig enhed med fokus på ”emerging markets”. Dernæst er der skabt en end-to-end løsning med relevante applikationer, som bruger data fra den virkelige verden for at samle feedback fra brugerne. Brugerens tilbagemeldinger analyseres i forbindelse med de ovennævnte udfordringer, som er typiske for udviklingslandene og omsættes til et sæt af krav.

Baseret på disse krav, introducerer rapporten et sæt nye lav-komplekse protokoller og sikkerhedsordninger, der adresserer den lave båndbredde / lave QoS netværks problemer og den lille regnekraft i enheden. Rapporten foreslår også en intelligent blanding af broadcast-tv og internet via kanal logo detektion og optisk tegngenkendelse for at håndtere en god brugervenlighed. Til sidst introducerer rapporten en ny tekstindtastning mulighed i fjernsynet ved hjælp af infrarøde fjernbetjeninger, der gennem en innovativ tastaturlayout på skærmen gør det brugervenligt for ikke computer kyndige brugere.

Acknowledgment

I am deeply indebted to my supervisor Prof. Ramjee Prasad, University of Aalborg for giving me the opportunity to work under his supervision. His guidance has really helped me in completing the work. I am also grateful to Prof. Prasad for letting me learn from him on not leaving things until finished. It was really great learning experience beyond technical fields.

I am grateful to my Professors at Indian Institute of technology, Kharagpur, Prof. R.V. Rajakumar and Prof. B.N. Chatterjee for making me believe that I would be able to work towards the thesis even when I am working. I also thank Prof. Anupam Basu from IIT, Kharagpur and Prof. Utpal Garain from Indian Statistical Institute, Kolkata for giving valuable inputs towards my work.

I thank TCS Chief Technology Officer Mr. K. Ananthakrishnan for allowing me to work part-time on my thesis; Mr. Debasis Bandyopadhyay, erstwhile Head of R&D for Ubiquitous Computing Theme in TCS for motivating and encouraging me to pursue my technical passions and providing me with all the support needed; and Dr. Sunil Sherlekar, erstwhile Head of Embedded System Group in TCS for helping in my initiation into Industry-focused R&D.

I am deeply thankful to all the co-authors in my publications and colleagues in TCS Innovation Lab Kolkata, especially Tanushyam Chattopadhyay, Aniruddha Sinha, Chirabrata Bhaumik, Avik Ghose, Debnarayan Kar, Soma Bandyopadhyay, M. Prashant, Jasma Shukla and Debatri Chatterjee among others, who have helped in the work.

I thank Prof. Albena Mihovska and Prof. Rasmus Hjorth Nielsen from CTIF, Aalborg University for their review and comments on the thesis. I also thank Susanne Nørrevang and Jens Erik Pedersen from CTIF, Aalborg University for helping me out with the administrative formalities.

Finally I must say that this work could not have been completed without the blessings from my parents, the encouragement and support from my wife Sanghamitra, and the last but not the least, the love and inspiration from my little daughter Tanisha.

Table of Contents

List of Acronyms vii
List of Figures ix
List of Tables xi
1. Introduction 1
1.1 Motivation 1
1.2 Challenges 2
1.3 Novelty and Contributions 3
1.4 Thesis Outline 4
References 5
2. Requirement Analysis 6
2.1 Introduction 6
2.2 Application Development Framework 7
2.2.1 Background Study 7
2.2.2 Framework Architecture 7
2.3 User Trial 9
2.3.1 Survey Configuration 9
2.3.2 Qualitative Survey 9
2.3.3 Quantitative Survey 9
2.4 Requirement Analysis 10
2.5 Conclusions 11
References 11
3. Video Chat over Low-QoS Networks 12
3.1 Introduction 12
3.2 Problem Definition 12
3.3 Proposed System 13
3.3.1 Sensing of Network Condition 14
3.3.2 Rate control in audio and video codecs 15
3.3.3 Adaptive packetization of the encoded data 16
3.4 Results 17
3.4.1 Experimental Setup 17
3.4.2 Experimental Results 17
3.4.3 Discussion 19
3.5 Conclusions 20
References 20
4. Low-complexity Video Security 21
4.1 Introduction 21
4.2 Low-Complexity Video Watermarking 21
4.2.1 Problem Definition 21
4.2.2 Proposed Watermarking Algorithm 23
4.2.3 Results 26
4.2.4 Discussion 31
4.3 Low Complexity Video Encryption 32
4.3.1 Problem Definition 32
4.3.2 Proposed Encryption Algorithm 32
4.3.3 Results 35
4.3.4 Discussion 36
4.4 Conclusion 36
References 37
5. Context-aware Intelligent TV-Internet Mash-ups 38

5.1 Introduction 38
5.2 TV Channel Identity as Context 39
5.2.1 Problem Definition 39
5.2.2 Proposed System 41
5.2.3 Results 42
5.2.4 Discussion 43
5.3 Textual Context from Static Pages in Broadcast TV 44
5.3.1 Problem Definition 44
5.3.2 Proposed System 44
5.3.3 Results 46
5.3.4 Discussion 46
5.4 Textual Context from Text Embedded in Broadcast TV 48
5.4.1 Problem Definition 48
5.4.2 Proposed System 49
5.4.3 Results 51
5.4.4 Discussion 51
5.5 Conclusion 53
References 54
6. Novel On-screen Keyboard 56
6.1 Introduction 56
6.2 Problem Definition 56
6.3 Proposed System 57
6.3.1 Proposed Algorithm for Optimal Layout 59
6.3.2 Implementation Details 60
6.4 User Study 61
6.4.1 Methodology 61
6.4.2 Results and discussion 62
6.5 Conclusion 64
References 65
7. Conclusion and Future Work 66
7.1 Conclusion 66
7.2 Future Work 68
Appendix A. Home Infotainment Platform 70
A.1 System Description 70
A.1.1 Hardware Details 70
A.2 Applications 71
A.2.1 Browser 71
A.2.2 Media Player 71
A.2.3 Video Chat 72
A.2.4 SMS on TV 72
A.2.5 Remote medical Consultation 73
A.2.6 Distance Education 73
A.3 Improvements in HIP on-screen keyboard 75
Appendix B. List of Publications and Patents 79

List of Acronyms

AA	Averaging Attack
AAC	Advanced Audio Coding
AAD	Average Absolute Difference
ADSL	Asymmetric Digital Subscriber Line
AMR	Adaptive Multi-Rate audio codec
API	Application Programming Interface
AVC	Advanced Video Coding
CAA	Circular Averaging Attack
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CPM	Characters per Minute
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
DRM	Digital Rights Management
DSP	Digital Signal Processor
DTH	Direct to Home
DTX	Discontinuous Transmission
DVB	Digital Video Broadcasting
ECG	Electrocardiography
EPG	Electronic Program Guide
FMO	Flexible Macro-block Ordering
FFA	Frequency Filtering Attack
FPS	Frames per Second
GA	Gaussian Attack
GCA	Gama Correction Attack
GOMS	Goal Operator Methods
GOP	Group of Pictures
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GSSNR	Global Sigma Signal to Noise Ratio
GUI	Graphical User Interface
HEA	Histogram Equalization Attack
HIP	Home Infotainment Platform
HSV	Hue, Saturation, Value
HTML	Hypertext Markup Language
HVS	Human Vision Psychology
ICT	Information and Communication Technology
IDR	Instantaneous Decoding Refresh
IPTV	Internet Protocol Television
IR	Infra-red
KLM	Keystroke-level-model
LEA	Laplacian Attack
LMSE	Laplacian Mean Square Error
MAD	Mean Absolute Difference
MB	Macro-block

MOS	Mean Opinion Score
MPEG	Motion Picture Experts Group
MSE	Mean Square Error
NALU	Network Abstraction Layer Unit
NLFA	Non-linear Filtering Attack
NLP	Natural Language Processing
NMSE	Normalized Mean Square Error
OCR	Optical Character Recognition
OTT	Over the Top
PAL	Phase Alternating Line
PC	Personal Computer
PCA	Principal Component Analysis
PPS	Picture Parameter Set
PSNR	Peak Signal to Noise Ratio
QA	Question-answer
QCIF	Quarter Common Intermediate Format
QOE	Quality of Experience
QOS	Quality of Service
QP	Quantization Parameter
RBSP	Raw Byte Sequence Payload
RF	Radio Frequency
ROA	Rotate Attack
ROI	Region of Interest
RSA	Resize Attack
RTP	Real-time Transport Protocol
RTT	Round-trip Time
SAD	Sum of Absolute Difference
SB	Sub Block
SDTV	Standard Definition Television
SIM	Subscriber Identity Module
SMS	Short Messaging Service
SNR	Signal to Noise Ratio
SPS	Sequence Parameter Set
TCP	Transmission Control Protocol
TS	Transport Stream
TV	Television
UDP	User Datagram Protocol
URL	Uniform Resource Locator
USB	Universal Serial Bus
VCD	Video Compact Disc
VGA	Video Graphics Array

List of Figures

Fig. 1.1: Mapping of Requirements to Technology Challenges 2
Fig. 1.2: Thesis Organization 4
Fig. 2.1: Framework Architecture for Different Applications 8
Fig. 2.2: Qualitative Feedback 9
Fig. 2.3: Quantitative Analysis – Ease of Use 10
Fig. 2.4: Quantitative Analysis – Likeability 10
Fig. 2.5: Quantitative Analysis – Relevance 10
Fig. 2.6: Quantitative Analysis – Text Entry and Navigation 10
Fig. 3.1: System Architecture for Rate Adaptive Video Chat 14
Fig. 3.2: Probe pair scheme for estimating bandwidth 14
Fig. 3.3: Quality Comparison - proposed algorithm vs. standard implementation 18
Fig. 3.4: Network Statistics in Different Scenarios 19
Fig. 3.5: Effective Bandwidth in Modem-ADSL network 19
Fig. 3.6: Effective Bandwidth in Modem-Modem network 19
Fig. 4.1: H.264 Encoder Architecture for Watermark Embedding 23
Fig. 4.2: Watermark Embedding Algorithm Flow Diagram 23
Fig. 4.3: Flow Diagram for Checking the Watermarking Message Size 24
Fig. 4.4: Location Selection for Embedding inside Image 24
Fig. 4.5: Algorithm Flow for Handling Data or Image for Watermarking 25
Fig. 4.6: Quality Degradation after Watermarking 27
Fig. 4.7: Architecture of Watermarking Evaluation Tool 27
Fig. 4.8: Snapshots of Video Frames and Retrieved Logo from Watermarked Video 29
Fig. 4.9: Block diagram of Read module 33
Fig. 4.10: NALU organization video conferencing application 33
Fig. 4.11: NALU organization for video storage application 33
Fig. 4.12: Flow chart of the process of encoding 34
Fig. 4.13: Modified FMO algorithm 34
Fig. 5.1: Using HIP for TV-Internet Mash-ups 39
Fig. 5.2: Channel Logos in Broadcast Video 39
Fig. 5.3: Logo Types 40
Fig. 5.4: Overview of Channel logo recognition 41
Fig. 5.5: Channel Logo Location Shift 43
Fig. 5.6: Channel Logo Color Change 43
Fig. 5.7: Channel Logo Image Change 43
Fig. 5.8: Text Embedded in Active Pages of DTH TV 45
Fig. 5.9: Text Recognition is Static Pages 45
Fig. 5.10: Different Active Page Screenshots (a) – (j) 47
Fig. 5.11: Performance of OCR Engines before and after the Proposed Algorithms 47
Fig. 5.12: Contextual Text Embedded in TV Video 48
Fig. 5.13: Text Recognition is Broadcast Video 50
Fig. 5.14: Screen shots showing breaking news in four different channels 52
Fig. 5.15: High Contrast Regions in the Video 52
Fig. 5.16: Text Regions after Noise Cleaning 52
Fig. 5.17: Screen shot of the Google Search Engine with Recognized Text as Input 53

Fig. 5.18: Screen shot of the Final Application with TV-Web Mash-up....53
Fig. 6.1: Layout of the Accompanying Remote Control 58
Fig. 6.2: Proposed Keyboard Layout - Lower case letters....58
Fig. 6.3: Algorithm Flowchart for Keyboard Layout Decision 59
Fig. 6.4: Proposed Layout - 1....60
Fig. 6.5: Proposed Layout – 2....60
Fig. 6.6: Traditional QWERTY Layout....60
Fig. 6.7: Improvement of Layout-1 over on-screen QWERTY Layout 63
Fig. 6.8: Improvement in Layout-1 after practice....63
Fig. A.1: Home Infotainment Platform System Block Diagram 71
Fig. A.2: Main Menu and Browser on HIP....72
Fig. A.3: Picture Viewer, Music Player and Video Player on HIP....72
Fig. A.4: Video Chat on HIP 72
Fig. A.5: SMS on HIP....73
Fig. A.6: Proposed System Architecture....74
Fig. A.7: Data Capture Screen 74
Fig. A.8: Data Upload Screen....74
Fig. A.9: Expert Doctor View....75
Fig. A.10: Video Chat Session....75
Fig. A.11: Solution Architecture – Satellite Broadcast....75
Fig. A.12: Solution Architecture – Internet based Interactivity....76
Fig. A.13: Lecture Video 76
Fig. A.14: Rhetoric Questions and Answer 76
Fig. A.15: Audio Question/Answer Recording and Playback 77
Fig. A.16: Proposed Keyboard Layout - Upper case letters 77
Fig. A.17: Proposed Keyboard Layout - Symbols....77
Fig. A.18: Updated Symbol and Smiley Layout....78
Fig. A.19: Visual effect of key-press and colored hot keys....78
Fig. A.20: Co-located help....78

List of Tables

Table 2.1: Framework Configuration for Different Applications....8
Table 2.2: Framework Mapping for Remote Medical Consultation....8
Table 2.3 Framework Mapping for Distance Education....8
Table 3.1 Effective Bandwidth Computation....14
Table 3.2: Network Statistics in Different Scenarios....18
Table 4.1: Theoretical Computational Complexity....26
Table 4.2 Measured Complexity....26
Table 4.3: Video Quality after Attacks....28
Table 4.4: Retrived Image Quality after Attacks....30
Table 4.5: Retrieved Text Quality after Attacks....30
Table 4.6: Overall decision making process and Guideline....31
Table 4.7: Result Summary for the proposed Algorithm under different attacks....31
Table 4.8: NALU Unit type....33
Table 4.9: Computational Complexity of the proposed algorithm....35
Table 4.10: Memory complexity of the proposed algorithm....35
Table 4.11: Video Quality Analysis....36
Table 5.1 : Confusion Matrix for Channel Logo Recognition....43
Table 5.2 Time complexity of Different Algorithm Components....43
Table 5.3: Raw Text Outputs from algorithms for different Active Pages....47
Table 6.1: KLM Operators....61
Table 6.2 : KLM-GOMS for Email Sending Application....62
Table 6.3: Basic Operations for KLM-GOMS....62
Table 6.4: User Study Results for Normal Text Entry....64
Table 6.5: User Study Results for Email Application – Time Measurement....64
Table 6.6: User Study Results for Email Application – Comparison of Layouts....64

1 Introduction

1.1 Motivation

The main motivation of this thesis has been to create a connected television solution for developing countries like India addressing the challenges and requirements of the mass market that can break the digital divide barrier through affordable and innovative approach.

According to Mark Weiser, who is known as the father of ubiquitous computing for his pioneering research work at Xerox Palo Alto Research Center (PARC) in 1988 [1], computers, "rather than being a tool through which we work", should "disappear from our awareness". Ubiquitous computing is about computers everywhere surrounding humans, communicating with each other and interacting with people in certain ways [2], [3], [4]. It is the third wave of the computing revolution, the first and second waves being mainframe computing, and personal computing. In the era of ubiquitous computing, there will be one person using many computers embedded in mobile phones, handheld devices, audio/video players, televisions (TV), watches, toasters, ovens and cars. In [5], one gets a nice overview of the ubiquitous campus use cases on smart kitchens, smart classrooms and smart wallets. In [2], Weiser introduces the concept of TV as the "casual computer" device.

As we embrace the ubiquitous computing technology, there is a visible trend across the world of moving from Personal Computer (PC) towards mobile phones, tablets and TVs as the preferred set of ubiquitous screens in our life. However, the scenario is a little different in developing countries like India. Market studies in India reveal some interesting facts. According to studies done by Indian Marketing Research Bureau (IMRB) [6], there were 87 Million PC literate people (out of 818 Million total population above age group of 12) and 63 Million internet users in 2009 in India. However, only 30% of these users accessed internet from home PCs. There were a sizeable 37% of users accessing the internet from cyber cafes and only 4% accessing from alternate devices like mobiles. More recent studies by International Telecommunication Union (ITU) [7] indicate that in 2010, household computer penetration in India was only 6.1% and household internet penetration was only 4.2%. This clearly brings out a clear picture of the digital divide that exists in India, where very little proportion of the population have access to PCs or Internet due to cost, skill, usability and other issues.

Reference [7] also indicates that mobile phone penetration in India was 61.4% in 2010. In another similar report [8], it is stated that India has about 812 Million mobile subscribers in 2011, however only 26.3 Million of them are active mobile internet users. This can be attributed to the fact that majority of the mobile phones in India are low-end having small size screens, thereby limiting the volume and quality of information that can be disseminated and the overall end-user experience. Tablets and large-screen smart phones, though having a larger screen size and a nice touch-screen

experience, are not yet available in an affordable price level. Similar kind of digital divide pictures emerge from other developing countries also [9], [10].

On the other hand, number of television sets used in India has reached more than 60% of homes (158 Million households in 2011[1]). In this context, if one could make the television connected to the internet world in a low-cost manner, it has the potential of becoming the "Ubiquitous Computing Screen" for the home helping in bringing down the above-mentioned digital divide because it already has high penetration and a large display screen capable of providing acceptable user experience.

However, the emerging markets like India are characterized by some unique challenges like low bandwidth / low Quality-of-Service (QoS) of the available wireless networks, extreme cost-consciousness of the users and lack of computer literacy among masses. Most of the available solutions are targeted towards high-end markets and none of them really address these technology challenges posed by the developing markets like affordability, poor network bandwidth/QoS, and usability for non-computer-savvy users etc. They also lack applications catering to the social and economic needs of masses in developing countries.

1.2 Challenges

Connected Televisions are already making their mark in the developed countries[2]. There are connected TV / Smart TV infotainment solutions from LG, Samsung, Vizio, Sony, Panasonic etc. [11], which tries to mix entertainment experience from TV with the information expereince from Internet. In [12], one gets specific details about Smart TV market potential in Korea and in [13], the concept of providing social applications on Smart TV in introduced. But none of these address the developing country specific challenges like low bandwidth / low Quality-of-Service (QoS) of the available wireless networks, extreme cost-consciousness of the users and lack of computer literacy among masses. Fig. 1.1 depicts the proposed scientific problems (in blue) in relation to the requirements (in green) and user level challenges (in red) mentioned in the previous section.

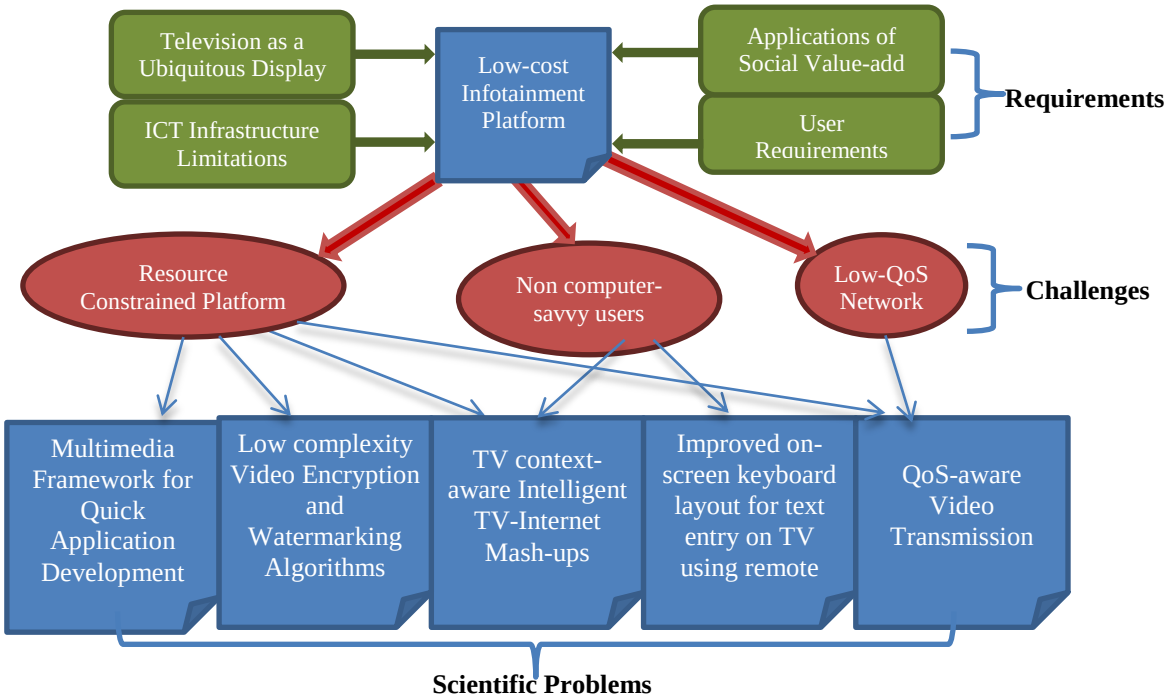


Fig. 1.1: Mapping of Requirements to Technology Challenges

As depicted in Fig. 1.1, the challenges stemming from market requirements translate into some unique scientific problems as listed below –

[1] http://en.wikipedia.org/wiki/Television_in_India

[2] http://www.informationweek.com/news/personal-tech/home-entertainment/219100136

1. Lack of software middleware and framework on low-cost processors like ARM, which tends to make the application development time and the overall system cost higher.
2. Video transmission is one of the most-affected applications, when the available bandwidth is low and QoS of the Wireless Access network fluctuates to a low level. There is need to design suitable video transport schemes to handle this.
3. Multimedia content sharing in an Infotainment system demands access control and digital-rights-management (DRM) support, but the encryption / watermarking algorithms required tend to have high computational complexity. In order to keep the system cost low by using low-cost constrained processing power platforms, there is need to have low computation complexity encryption and watermarking algorithms.
4. The lack of computer literacy among masses demands the browsing experience to seamlessly blend into the broadcast TV viewing experience. This in turn requires intelligent mash-ups of broadcast TV and internet content with low computational complexity.
5. To keep the system cost low and due to lack of computer-literacy of the target users, it is required that the system is operated by the users using a familiar input device like an infra-red remote control. This in turn poses a unique challenge of how to provide acceptable text-entry experience on TV using remote control and on-screen keyboard.

The above scientific problems are addressed in this thesis to come up with a set of contributions leading towards an end-to-end system solution.

1.3 Novelty and Contributions

As the engineering contribution towards the business aspect of addressing the digital divide problem, an affordable over-the-top box (called Home Infotainment Platform or HIP) is proposed that can connect to internet and has TV as the display. To keep the box cost low, it uses a low-cost low-computing power processor (ARM) with multimedia accelerators and with a lot of open-source software and value-added applications. The system is already built and deployed in pilot scale in India and Philippines.

As the first scientific contribution a novel multimedia application framework is proposed to be used as the initial reference framework for carrying out the requirements analysis. The application framework on the ARM CPU based low-computing-power platform of HIP is tightly integrated to the DSP accelerator-core available on the CPU. This is advantageous for quick application development on the HIP, thereby reducing effort, time and cost. HIP is used as a platform for conducting field trials and user studies. As a subsequent (second) contribution, an in-depth analysis of the user study results is provided to ratify the problem requirements leading towards further scientific contributions in the thesis as listed below.

The third contribution in the thesis is an enhanced version of existing video transport protocols that ensure performance in unreliable low-QoS network conditions. The proposed enhancements relate to bandwidth sensing, rate-adaptation and packet fragmentation to provide an end-to-end system design approach.

The fourth contribution is a set of novel low computational complexity compressed-domain encryption and watermarking schemes for video that can be implemented in real-time on the low-processing power CPU of the box. The thesis also introduces novel methodologies to evaluate watermarking and encryption performance from security angle without compromising on video quality.

The next contribution is a novel TV-internet mash-up system that recognizes the broadcast TV context through channel logo detection and optical character recognition and then provides context-relevant information from internet in form of mash-ups and blending. As specific contributions, it proposes low-computational complexity channel logo detection algorithms; improved pre-processing, text localization and post-processing algorithms for text recognition in videos.

The final contribution of the thesis is an improved text-entry scheme using infra-red remote controls through novel layout design of an on-screen keyboard, that will help the end-users operate the system in a lean-back mode as they are used to for viewing traditional TVs. The proposed on-screen keyboard layout significantly reduces the typing load for doing text entry. As an additional contribution, the thesis also introduces an improved keyboard layout design space exploration methodology based on user study.

All the contributions have been published peer-reviewed journals and conferences. The work has resulted in 9 conference publications ((Appendix B - [1] to [9]), 3 book chapters (Appendix B - [10], [12], [13]) and one journal paper (Appendix B - [15]). Further two journal publications are in submitted stage (Appendix B - [11], [14]). Most of the ideas presented here are also filed for patents and they are validated for their novelty, non-obviousness and utility during the patent search process. There are 8 patent filings in total, of which 1 patent is already granted. The relevant patents are listed in Appendix B ([a] to [h]).

1.4 Thesis Outline

The thesis is organized in form of a monograph with 7 chapters outlining complete details of the work. All the relevant own publications are referenced in Appendix B. With reference to Fig. 1.1, the thesis organization is described in Fig. 1.2.

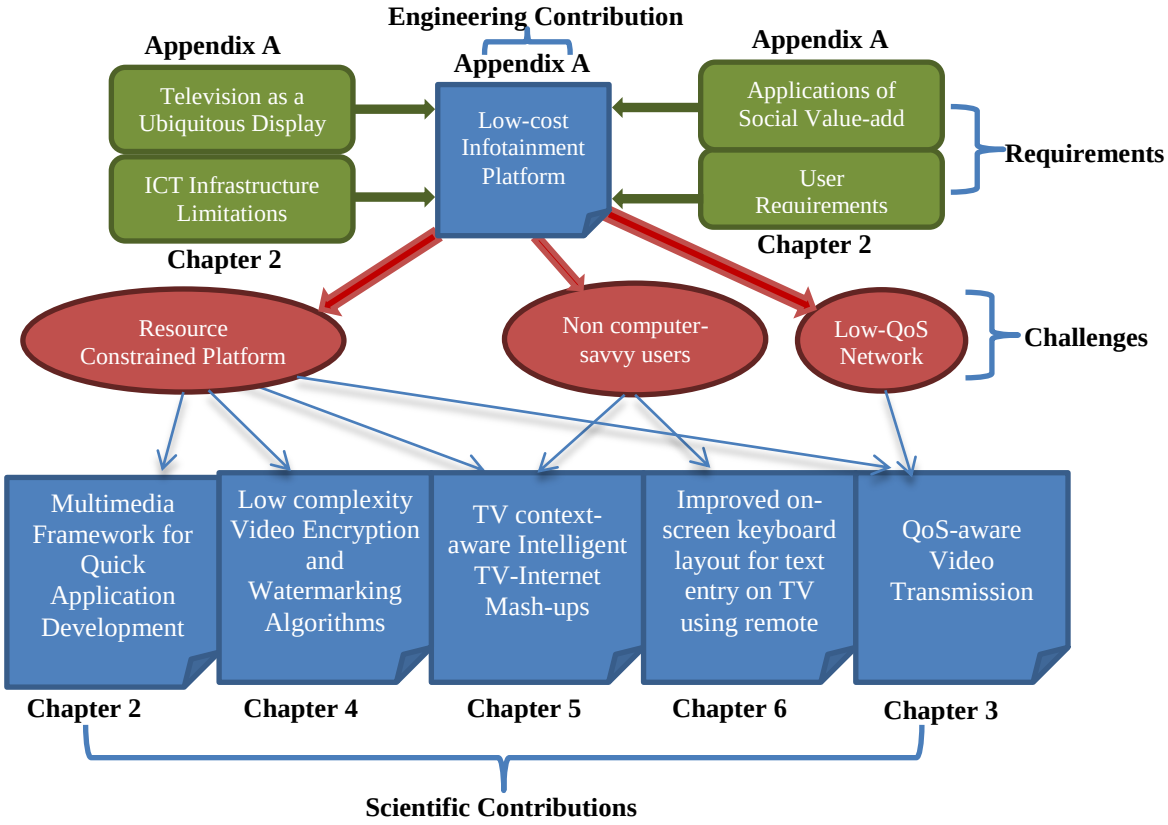


Fig. 1.2: Thesis Organization

Chapter 1 (current chapter) is the Introduction section which discusses about motivation of using television as a ubiquitous computing device, and maps user level requirements of developing countries to specific technology challenges leading to the novel contributions for the thesis.

Chapter 2 describes the user and technical requirements analysis based on which the scientific contributions have been made. For the purpose, a novel application development framework on top of a low-cost over-the-top Box (called Home Infotainment Platform or HIP) has been designed as the reference framework. It then presents the results of a user study conducted on HIP and analyzes it to arrive at the problem requirements for this thesis. All the subsequent work presented in the thesis is

derived from these requirements. The work on the novel framework on HIP is already published (Appendix B - [1], [10]). One journal paper (Appendix B – [11]) is also published on the user study based Requirement Analysis.

Chapter 3 proposes a novel bandwidth-aware, rate and fragmentation adaptive video chat solution that provides an acceptable video chat quality even in adverse network scenarios and presents results on real network conditions. The work has already been published (Appendix B - [2]).

Chapter 4 introduces novel low-computational-complexity encryption and watermarking schemes embedded inside the video encoder / decoder that can be run in real-time on the low speed CPU of HIP for secure video chat and video content sharing applications. It also provides a methodology to evaluate the security performance of the encryption and watermarking schemes. These works on watermarking and encryption have been published (Appendix B - [3], [4], [12]).

Chapter 5 proposes an implementation for the Smart TV platform built on top of HIP that can provide novel value-added context aware applications on television that seamlessly blends Internet content with broadcast TV content. It presents computationally efficient and improved accuracy algorithms for detecting television context through channel logo detection and embedded text recognition in TV videos. All these work have resulted in publications (Appendix B - [5], [6], [7], [13]). One journal paper (Appendix B – [14]) is in submitted state.

Chapter 6 introduces a novel on-screen keyboard layout optimized for infra-red remote controls that can be used to improve human-computer interaction on the HIP. A design space exploration based evaluation methodology using user study results is also proposed for evaluating the efficacy of the proposed layout. The works on both the layout and design space exploration has been published (Appendix B ([8], [9], [15]).

Chapter 7 summarizes the work done and outlines the future scope of work.

Appendix A provides description and configuration of the Home Infotainment Platform (HIP and elaborates on the applications on top of it.

Appendix B provides a list of own publications and patents.

References

[1]. M. Weiser, “Some computer science issues in ubiquitous computing”, *Commun. ACM*, 36(7):75-84, 1993.

[2]. Mark Weiser, "The world is not a desktop", *Interactions*; January 1994.

[3]. Mark Weiser, "Hot Topics: Ubiquitous Computing", *IEEE Computer*, October 1993.

[4]. M. Weiser, “The computer for the 21st century”, *Scientific American*, September 1991.

[5]. M. Weiser, “The future of ubiquitous computing on campus”, *Commun. ACM*, 41(1), January 1998.

[6]. Indian Market Research Bureau, “I-Cube 2009-10”, February 2010.

[7]. International Telecommunication Union (ITU), “Measuring the Information Society”, 2011.

[8]. Internet and Mobile Association of India, Report on “Mobile Internet in India”, Aug 2011.

[9]. International Telecommunication Union (ITU), “The World in 2011 – ICT Facts and Figures”, 2011.

[10]. International Telecommunication Union (ITU), “Information Society Statistical Profiles – Asia and the Pacific”, 2011.

[11]. Patrick Miller, "The 5 Best Smart TV Platforms of 2011”, *PC World*, Sept. 2011.

[12]. MoonKoo Kim and JongHyun Park, “Demand Forecasting and Strategies for the Successfully Deployment of the Smart TV in Korea”, *13th International Conference on Advanced Communication Technology (ICACT),* Feb. 2011.

[13]. Kwihoon Kim et. al., “Research of Social TV Service Technology based on SmartTV platform in Next Generation Infrastructure”, *5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT),* Dec. 2010.

2
Requirement Analysis

2.1 Introduction

There are three screens in the average user's life – personal computer / laptop / tablet, television and mobile phone / smart phone. Looking from the India and other developing countries perspective, personal computer / laptop / tablet are still not quite affordable to masses. Personal computer / laptop also suffer from usability problems for non-techno-savvy users. Mobile phone, though cheap and affordable, suffers from its very small display screen real-estate, which prevents detailed information dissemination and rendering on the screen. Smart phone and tablet do have larger screens but their cost is quite high for mass acceptance.

On the contrary, analog Television is a pervasive device that has invaded most of the homes in the developing countries (as outlined in section 1.1) and if this can be turned into an infotainment device through some low-cost add-on, it can have the following three advantages –

- Large-screen real estate to disseminate and render rich information
- Low incremental cost of ownership, thereby being affordable to the masses
- Simple user interface using the familiar TV-like remote control, thereby addressing the non-techno-savvy user needs

With this idea in mind, a low-cost add-on device to analog television called Home Infotainment Platform (HIP) had been created. This is described in detail in Appendix A. The basic version of HIP has already been deployed in a pilot scale in two countries – India and Philippines. In India, it was launched through the telecom service provider Tata Teleservices under the brand name of "Dialog"[3]. In Philippines, it was launched through the telecom service provider Smart Telecom under the brand name of "SmartBro SurfTV"[4] [5].

HIP is used as a platform to perform user studies in real deployment scenarios and understand end-user requirements for developing countries like India. As one of the contributions in this chapter, a flexible application development framework on HIP is proposed to quickly prototype applications required for performing the user study. The user study is performed with a group of actual consumers in India using a set of applications on HIP outlined in Appendix A. As the main contribution in this chapter, analysis of the results of the user study is presented that reveals a set of scientific challenges relevant to developing countries. Based on this analysis the problem requirements are formulated for

[3] http://telecomtalk.info/ttsl-intros-dialog-now-access-internet-over-television-with-tata-photon/20919/
[4] http://www.reviewstream.com/reviews/?p=102180
[5] http://www.gadgetpilipinas.net/2010/03/smart-releases-smartbro-surf-tv-sets-doomsday-to-internet-shop-owners/

addressing them in the subsequent chapters. The work on the novel framework on HIP is already published (Appendix B - [1] and [10]). The novel contributions of the work are also filed for a patent (Appendix B – [a]).

In section 2.2 the design of the novel application development framework is presented to show how the applications outlined in Appendix A can be mapped into the framework for quick prototyping. In section 2.3 the results of the user trial conducted in India through the "Dialog" pilot pre-launch is presented and in section 2.4, the results are analyzed to arrive at the problem requirements for addressing in subsequent sections. Finally the chapter ends with a conclusion section (2.5).

2.2 Application Development Framework

2.2.1 Background Study

The proposed framework is somewhat similar to mobile operating systems like Android and iOS, which builds on top of a core operating system kernel (mostly linux / unix) and provides APIs to build applications. However all such mobile operating systems are designed for small screen size and hence have a performance problem if the display resolution is increased to support large screens like TV, especially if the CPU has low processing power. Additionally, iOS is a closed system and hence cannot be deployed on any custom hardware. There is some framework described in [1], however it focuses mainly on IPTV. Intel and Nokia supported Meego[6] has been the closest match for Connected TV environment, but they never came out with a TV focused release addressing the above problems. There are also multimedia-specific solutions like Boxee[7] which was initially built for converting PCs into TVs through streaming media framework and then was converted for Set top boxes. It mostly supports all the features required in the proposed system, however, it is supported only on Intel x86 architecture based CPUs like ATOM, which are costly.

To keep the box cost low, it was decided to go for an ARM based CPU with multimedia accelerators and hence the complete framework had to be designed and developed from scratch. The system architecture describing the framework is given below.

2.2.2 Framework Architecture

The proposed architecture is designed create a scalable and flexible application development framework on top of HIP (Appendix A). Fig. 2.1 shows the details of the framework. It consists of three distinct but closely knit subsystems –

1. The source subsystem (SRC) - defines from where the data has to be taken for processing
2. The processing subsystem (PROC) - defines the kind of processing that needs to be done on the data to make it suitable for output
3. Finally the sink subsystem (SINK) - defines where the data has to be put after it is processed

SRC, PROC and SINK consist of the following modules in HIP (details in Appendix A) -

SRC – Network, Microphone, Camera, Audio in, Video in, Storage, USB
PROC – Compress / Decompress, Multiplex / Demultiplex, Render, Blend
SINK – Network, VGA, TV Video, TV Audio, Headphone, Storage

There is a control sub-system which can set which modules from SRC, PROC and SINK should be used and can control the data flow between chosen modules. It exposes Application Programing Interfaces (APIs) for applications to configure the settings as per their requirements. The framework

[6] www.meego.com
[7] www.boxee.tv

works for both multimedia and sensor data. Table 2.1 shows how choosing specific modules for SRC, PROC and SINK can create different applications. As seen from the table, it is clear that the proposed framework is flexible enough to support multitude of applications. Table 2.2 and Table 2.3 show how additional applications in medical and education space described in Appendix A can be mapped into the framework.

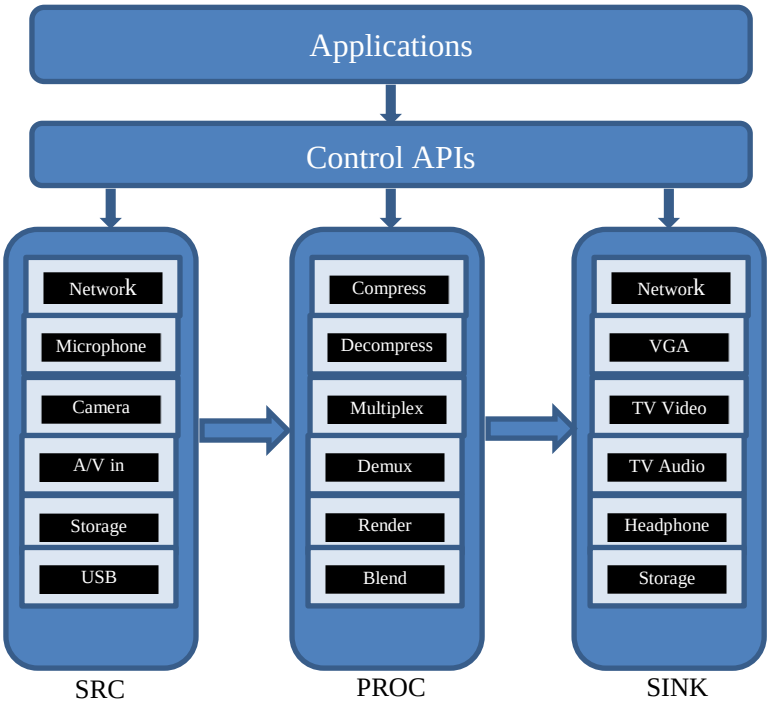


Fig. 2.1: Framework Architecture for Different Applications

Table 2.1: Framework Configuration for Different Applications

Application	SRC	PROC	SINK
Video from Internet	Network	Demux – Decompress	TV Video / Audio
Media player	Storage	Demux – Decompress	TV Video / Audio
Video Chat (Far View)	Network	Demux - Decompress	TV Video / Headphone
Video Chat (Near View)	Camera and Microphone	Compress – Multiplex	Network
TV Video Recording	A/V in	Compress – Multiplex	Storage
Internet Browser	Network	Render	TV Video
TV-Internet Mash-up	Network and A/V in	Render – Blend	TV Video

Table 2.2: Framework Mapping for Remote Medical Consultation

Application Use Case	SRC	PROC	SINK
Medical Sensor Data	USB	Compress - Multiplex	Network
Media player	Storage	Demux - Decompress	TV Video / Audio
Video Chat (Far View)	Network	Demux - Decompress	TV Video / Headphone
Video Chat (Near View)	Camera and Microphone	Compress - Multiplex	Network

Table 2.3 Framework Mapping for Distance Education

Application Use Case	SRC	PROC	SINK
Guide for Lecture Content	Network	Demux Parser	TV Video / Audio
Lecture Recording	A/V in	Demux – Decompress, Remux - Compress	Storage
Lecture Playback	Storage	Demux – Decompress	TV Video / Audio
Rhetoric Question	Network	Demux Parser	TV Video / Audio
Audio Question Record	Microphone	Mux - Compress	Network
Audio Answer Playback	Network	Demux – Decompress	TV Audio

2.3 User Trial

2.3.1 Survey Configuration

A user trial survey with HIP (Pilot version of "Dialog" launched by Tata Teleservices Ltd.) was conducted among the city users in India [2]. The applications provided were Browser, Media Player (Photo, Music, and Video), SMS and Simultaneous viewing of TV and Browser through blending. The sample taken was 50 middle-class and lower middle-class families involving 50 working adults and 50 students (12-18 years age).

2.3.2 Qualitative Survey

A detailed questionnaire based survey was undertaken to gather the user feedback on their experience on using HIP. At first there were some qualitative questions on which respondents were asked to answer. The results are shown in Fig. 2.2 in terms of % of respondents answering positive to the question stated. As seen from the results, slow internet connection and dislike of TV-Browser blending had been the biggest concern. There was also significant and distinct difference in opinion between adults and students for likeability of different applications.

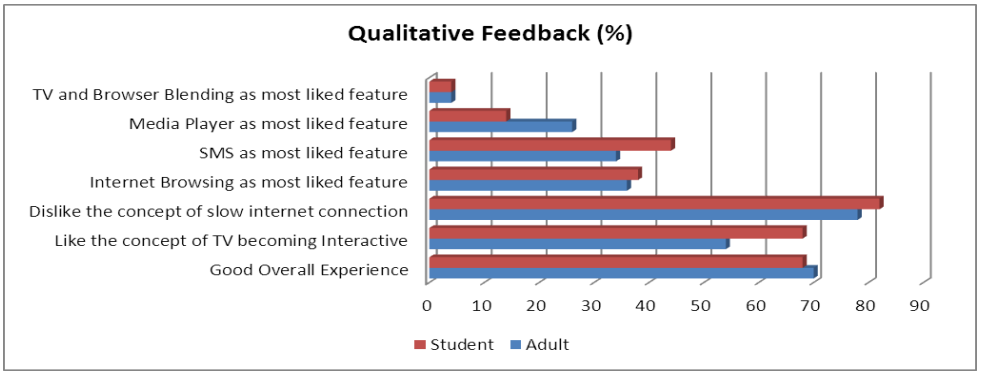


Fig. 2.2: Qualitative Feedback

2.3.3 Quantitative Survey

To gain more insight into the user feedback, a detailed quantitative study was also undertaken where users were asked questions and asked to rate each application (in a scale of 5) in three different dimensions – ease of use, likeability and relevance. The results are summarized using a confidence score measure where % of respondents responding with a score of 4 or 5 is given in the x-axis for each application. The y-axis lists the applications. These are given in Fig. 2.3, Fig. 2.4 and Fig. 2.5. As seen from these results, the striking points in the user feedback are lack of ease of use in internet browser; low likeability and relevance of photo viewer; significant difference in opinion among adults and students for internet browser and SMS and consistent below average rating for the TV-browser blending. Interestingly these findings from the quantitative study match closely with the findings of the qualitative study outlined in previous section.

Finally another quantitative study was undertaken to understand the navigation and text-entry issues. The results are summarized using a confidence score measure where % of respondents responding with 4 or 5 is given in the x-axis for different text-entry methods and is plotted in Fig. 2.6. The y-axis lists the different keyboards under test. As seen from the results, external keyboard is always a better option but it involves extra cost that reduces the mass-market applicability of the solution. Keeping aside the external keyboard results, it is seen that the adults prefer remote control as a preferred navigation device compared to keyboard. This can be linked to them being more non-computer-savvy. It is also seen that the comfort level for using the standard QWERTY on-screen keyboard layout is not good at all.

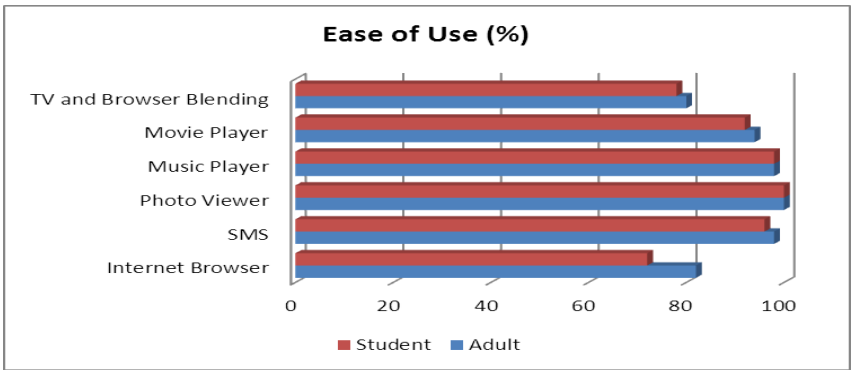


Fig. 2.3: Quantitative Analysis – Ease of Use

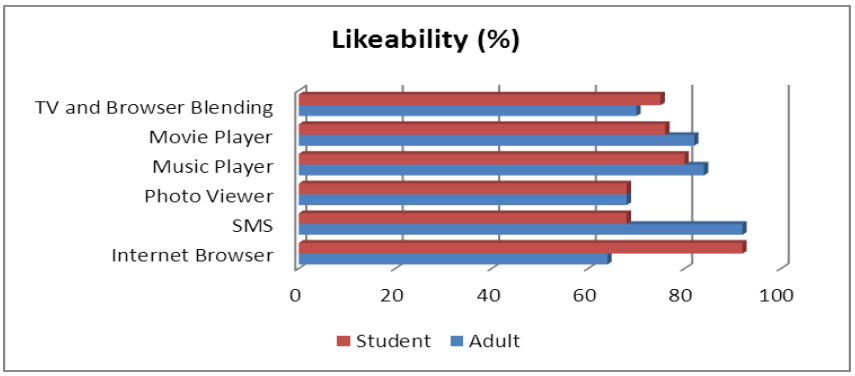


Fig. 2.4: Quantitative Analysis – Likeability

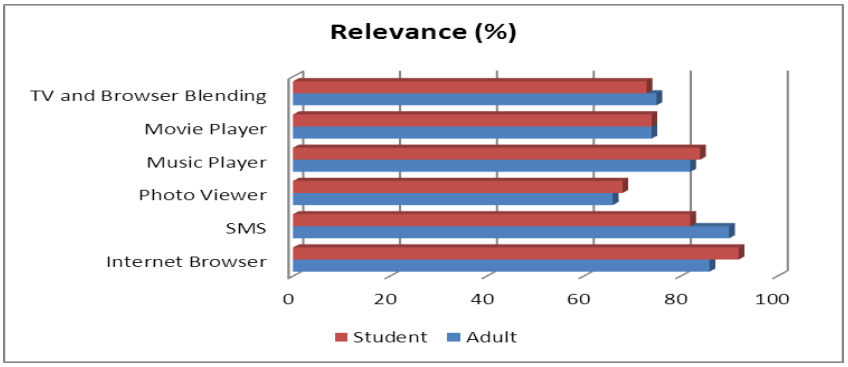


Fig. 2.5: Quantitative Analysis – Relevance

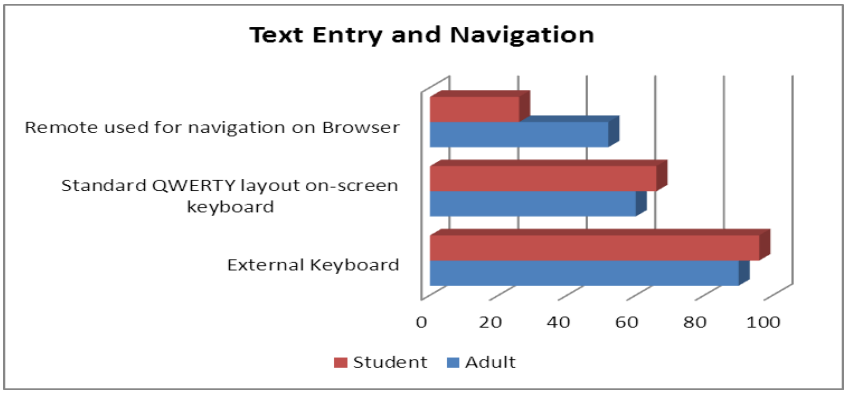


Fig. 2.6: Quantitative Analysis – Text Entry and Navigation

2.4 Requirement Analysis

Analyzing the findings of the user study, one can identify the following points of concern –

1. Slow to very-slow internet connection affecting the user experience.

2. Different levels of feedback from adults and students especially for browser and SMS, which can be attributed to generation specific preferences.
3. Poor feedback on Photo Viewer which on further analysis can be attributed to the poor Graphical User Interface (GUI) design.
4. Lack of ease of use for Internet Browser and not much liking for TV-Internet blending
5. Preference of remote control for non-computer-savvy users and dislike of the QWERTY layout for on-screen keyboard.

Keeping aside points 2 and 3 which point more to marketing and engineering challenges, focus is given on points 1, 4 and 5 which reveal some interesting scientific problems.

Regarding point 1, there is nothing can be done to improve the slow internet connection as it is an infrastructural problem. However it needs to be addressed from user experience perspective especially for bandwidth hungry applications like video streaming. Video chat was not even included the questionnaires for the user study, as it was seen during testing that it gave a very poor user experience due to poor QoS internet connectivity. This very specific QoS issue for Video Chat is addressed in **Chapter 3**.

Point 4 reveals an interesting user behavior – they do not want to watch TV and browse internet simultaneously as this is probably distracting. Also, people find using the browser on TV using remote to be difficult. These issues are addressed in **Chapter 5** through novel context-aware TV-Internet mash-ups.

Point 5 unearths the need for a remote based on-screen keyboard that is better than standard QWERTY layout –this is addressed in **Chapter 6** through a novel on-screen keyboard layout design.

Additionally, the other two socially value-adding applications available on HIP in healthcare and education (**Appendix A**), were not covered in the user study due to the need of other support in form of expert doctors, teachers etc. However, one concern that came up while discussing these applications requirements was the access control and rights management of the multimedia content, whether it is education content or video chat between patient and the doctor. This concern is addressed in **Chapter 3** through novel low-complexity real-time video encryption and watermarking systems.

2.5 Conclusions

In this chapter a novel application development framework on top of a low-cost connected TV solution called Home Infotainment Platform (HIP) was introduced. The proposed framework is flexible enough to deploy multitude of infotainment applications. It was used to develop different infotainment applications on HIP and its mapping to a wide variety of applications was presented to prove its flexibility and scalability. The system is already deployed in pilot scale in India and Philippines through two leading telecom operators in the two countries. Finally results of a user study on HIP based on real-world pilot deployment in India is presented and analyzed to arrive at the problem requirements. Based on this analysis, the technology challenges are firmed up in the form of acceptable video chat quality over low-QoS connectivity, secure and rights protected multimedia content delivery, efficient context-aware TV-Internet mash-ups and easy-to-use text entry methods using remote controls. These challenges are addressed in subsequent chapters.

References

[1]. ITU-T Technical Paper on "Delivery and control protocols handled by IPTV terminal devices", *SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS Infrastructure of audiovisual services – Communication procedures,* March 2011.

[2]. Market Study Report of Home Internet Product – *Tata internal.*

3
Video Chat over Low-QoS Networks

3.1 Introduction

Video chat and other video streaming applications are not only bandwidth hungry, but also have strict real-time packet delivery requirement. Since wireless networks have fluctuating conditions due to fading and other issues, maintaining the QoS of such applications over wireless network pose a big challenge. This especially true for low-QoS 2G mobile wireless data networks prevalent in developing countries like India.

In this chapter, a novel adaptive rate control for the video chat implemented on HIP is presented. It combines instantaneous sensing of the network conditions with bit-rate control of the audio/video codec and adaptive packet fragmentation to achieve better quality of experience for the end user. The solution can only be applied in scenarios where both client and server can implement the proposed algorithms, as is the case for video chat. Additionally, the instantaneous sensing of the network conditions can also be used as a network condition indicator for any bandwidth intensive operations with implementation control only on the client side (e.g. internet video streaming, multimedia-heavy browsing etc.).

In section 3.2, the problem statement is defined with the help of an application use case followed by a gap analysis through state-of-the-art study. In section 3.3, a novel rate adaptive video chat system is proposed that adapts in terms of compression quality (bit-budget), frame-rate, resolution and fragmentation after pre-sensing the network conditions. In section 3.4, the experimental setup and results are presented to show how the proposed system improves the performance. Finally in section 3.5 a conclusion and summary is provided.

3.2 Problem Definition

In countries like India, for remote rural areas and small cities, 2G wireless networks like GPRS and CDMA 1xRTT are the prevalent means of internet access because reach of the wire-line broadband is limited and 3G wireless is still very costly and not widely deployed. In such network conditions, it is often noticed that the data sent over the wireless link experience losses and variable amounts of delay. This creates difficulties in developing application like video chat that requires playing audio and video on real time that in turn require timely and in-order arrival of data packets to the destination. So there is a need for some mechanism at the transmitter end so that the data rate can vary adaptively with the available network capacity thereby maintaining a tradeoff between video quality and bitrate, with an additional constraint that audio should not be disrupted at all.

Some literature is available with different system approaches for the video conferencing solution for low bandwidth scenario. In [1] a four way video conferencing solution is discussed but there is no focus on the constrained bandwidth scenario. Flora[8] presents a low bandwidth video conferencing solution, but uses gesture detection to reduce the bit rate rather than sensing the network condition – similar approach is also taken in [2]. In [3], authors discuss about wireless network aware rate adaptive video streaming, but focuses only on WiFi.

Quite a bit of work has been done in the network condition estimation. In [4] there is a good comparative study of bandwidth estimation tools, but these tools cannot be implemented in a constrained platform like HIP. Karthik et. al., in [5] provides bandwidth estimation tools for WiFi and Cable Modem networks, but not for 2G cellular networks. In [6], Bolot et. al. uses the measured round trip delays of small UDP probe packets sent at regular time intervals to analyze the end-to-end packet delay and loss behavior in the Internet. In [7], Mascollo et. al. provides a mechanism to estimate network congestion in IP networks, but would need a component running on the server side, which makes it difficult to deploy in a service provider agnostic way. In [8], Keshav et. al. discuses about achieving and maintaining data transmission rates independent of the communication channel and includes techniques for data transmission initialization, data retransmission, and buffer management. In [9], Matta et. al. provides a methodology for QoS determination. However none of these papers look at the problem from end-to-end perspective and say how the estimated bandwidth be effectively and optimally used for audio/video streaming.

On the other hand, rate control from image processing and content analysis perspective is well researched. Reference [10] gives the rate control methodology in H.264. [11] and [12] also provides different ways for rate control in Video compression using the video content as driver. But none of them look at driving the rate control from available network capacity and low computational complexity perspective.

The main contribution of the current work is to propose an end-to-end video chat solution that senses the available bandwidth of a fluctuating GPRS or CDMA 1xRTT wireless channel and automatically adapts the video chat compression quality in an efficient and integrated way. As part of scientific contribution towards the solution, the following is proposed - an experimental heuristics based mapping of effective bandwidth to probe packet delay, a low complexity video rate control algorithm through automatic switching between frame and macro-blocks and an adaptive scheme for audio/video packet fragmentation. The work has been published in Appendix B - [2]. A patent is also filed on the same (Appendix B – [b]).

3.3 Proposed System

The system architecture for the proposed adaptive rate control based video chat solution is given in Fig. 3.1. It is implemented using the framework introduced in **Chapter 2**. Referring to Fig. 3.1, the encode / decode control belong to the control subsystem, audio/video encode / decode blocks belong to the processing subsystem and probe packet, packetization / depacketization, transmission / reception blocks belong to Network subsystem. The dark colored blocks in Fig. 3.1 denote the proposed enhancements.

The proposed system uses AMR-NB codec for audio and H.264 baseline profile for video as they provide the maximum compression for video chat type quality. It mainly has three components.

- Sensing of Network Condition
- Adaptive rate control in audio and video codecs based on sensed network condition
- Adaptive packetization and transmission interval of the encoded data based on sensed network condition

The design of each of these components is given in detail below.

[8] http://people.cs.uct.ac.za/~mnoordie/tmvumbi/lit_rev_flora.pdf

3.3.1 Sensing of Network Condition

Of all the network condition estimation systems discussed in section 3.2, ([4], [5], [6], [7], [8], [9]), it was found that probe-pair packet based algorithms give acceptable estimation accuracy while being computationally light. A similar mechanism is followed in the proposed system. The probe-pair packets are transmitted by the transmission end and returned back by the receiver end. The details of transmission of probe packets are shown in Fig. 3.2.

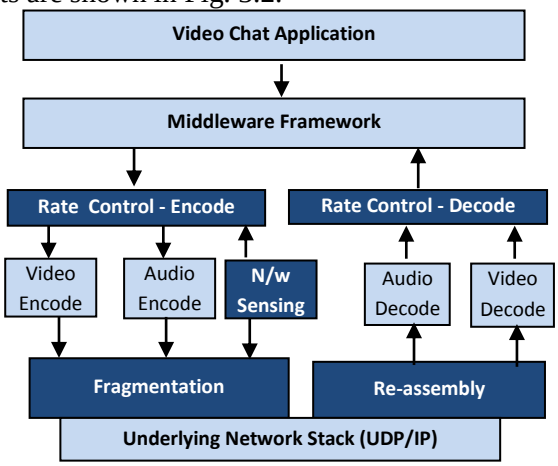


Fig. 3.1: System Architecture for Rate Adaptive Video Chat

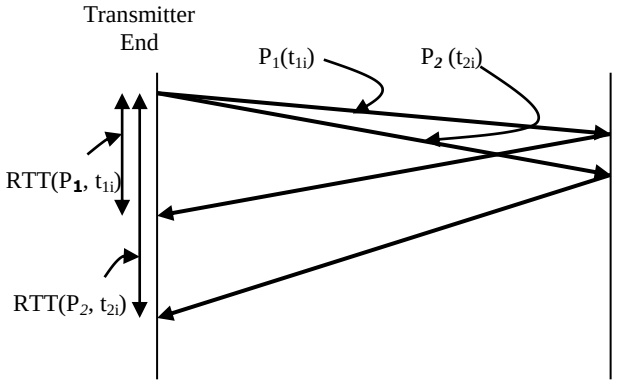


Fig. 3.2: Probe pair scheme for estimating bandwidth

Each time, two probe packets are transmitted $P1(t_{1i})$ and $P2(t_{2i})$. The time interval T between sending probe-pair packets is a function of the RTT of the probe packets can be computed as

$$T = average\ (RTT\ (P1, t_{1i}), RTT\ (P2, t_{2i})) \quad \textbf{(3.1)}$$

where,

$RTT\ (P1, t_{1i})$: the round trip time of the first probe packet transmitted at time t_{1i},

$RTT\ (P2, t_{2i})$: the round trip time of the second probe packet transmitted at time t_{2i}.

It should be noted that the two packets are sent just one after another and hence difference between their transmission times can be taken to have negligible effect. The effective bandwidth of the network BW_{eff} can be taken as inversely proportional to T. A heuristic mapping of BW_{eff} to T is found out through experimentation with large number of packets on the real network and is given in Table 3.1.

Table 3.1 Effective Bandwidth Computation

T (msec)	BW_{eff} (kbps)
T < 300	50
300 < T < 800	13
800 < T < 1600	4
1600 < T < 1900	2
T > 1900	1

3.3.2 Rate control in audio and video codecs

AUDIO CODEC

The AMR-NB speech codec [13] has an inherent support for eight different bit rates ranging from 4.75 kbps to 12.2 kbps having options for 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbps. For the video chat application on HIP, default rate was taken as 5.15 kbps and when the network condition became bad (BW_{eff} < 4 kbps), bit rate was changed to 4.75 kbps, the lowest option available.

VIDEO CODEC

H.264 based video compression [14] is used, as it is found to be most efficient in terms of bandwidth. According to [10], quantization parameter may vary according to the basic unit size in frame level, slice level or macro-block level. Larger basic unit size increase bit fluctuation and reduces time complexity. On the other hand, smaller basic unit increases time complexity but reduces bit fluctuation. Here an algorithm is proposed where the basic unit size is determined adaptively depending upon complexity of the scene and available bandwidth (BW_{eff}).

Image quality of any video sequence is usually measured in terms of Peak-Signal-to-Noise ratio (PSNR). It is found that PSNR increases with number of bits consumed in encoding. Thus image quality improvement in a constrained network condition can be thought of as a balancing act between of bit-budget and PSNR. Image quality is normally measured in terms of rate-distortion curve where PSNR is plotted against number of bits used in encoded stream. Rate distortion can be minimized by using rate distortion optimization (RDO) model [13], which is highly computationally expensive and not suitable to meet real time criteria in an embedded implementation like HIP. In this section first all the assumptions made are stated, followed by discussion on the state of the art of rate control and finally the proposed algorithm is presented. The advantage of the proposed algorithm is shown by comparing the PSNR value obtained by the proposed algorithm and the H.264 reference code for same bitrate. Assumptions for implementation are as follows:

- To reduce complexity, each frame is considered as a slice.
- The proposed approach should be generic enough to be applicable to any H.264 encoder environment.

In order to achieve constant bit rate (CBR), one needs to compute quantization parameter (qp) depending upon available bits and Mean absolute difference (MAD) of original and predicted image. Thus bitrate *b* may be represented as:

$$b = f(qp, MAD) \; where \; 0 \le qp \le 51 \tag{3.2}$$

Rate control may be achieved only by manipulating qp depending upon MAD and H.264 standard suggests that qp may vary in macro-block layer, slice layer or frame layer as delta qp is specified in all these layer headers. So this rate controlling can be performed in group of picture (GOP) level, frame level, slice level or macro-block (MB) level. Each of these layers is treated as the basic unit for rate controlling. Theoretically, basic unit in rate controlling is defined to be a group of continuous MBs.

But MAD can be computed only after reconstruction is done which in turn requires qp value. Thus it is something like the chicken and egg problem. To tackle this problem, a linear model is used to predict the MAD of the remaining basic units in the current frame by using those of the co-located basic units in the previous frame. Suppose that the predicted MAD of the l^{th} basic unit in the current frame and the actual MAD of the l^{th} basic unit in the previous frame are denoted by $MAD_{cb}(l)$ and $MAD_{pb}(l)$, respectively. The linear prediction model is then given by

$$MAD_{cb}(l) = a_1 \times MAD_{pb}(l) + a_2 \tag{3.3}$$

The initial value of a_1 and a_2 are set to 1 and 0, respectively. They are updated by a linear regression method similar to that for the quadratic R-D model parameters estimation in MPEG-4 rate control after coding each basic unit. So now the problem of image quality enhancement boils down to detecting the *qp* for a basic unit depending upon predicted MAD and target bits which in turn is computed from BW_{eff}. But it is also noted that by employing a big basic unit, a high PSNR can be achieved while the bit fluctuation is also big. On the other hand, by using a small basic unit, the bit fluctuation is less severe, but with slight loss in PSNR. Over and above, smaller basic unit increase the computational cost. So it is better to perform rate controlling with frame as basic unit when MB is not complex and with MB as basic unit when that frame contains complex picture information. The novel contribution of this work lies in selecting basic unit adaptively depending upon MAD, which optimizes bit distortion as well as time complexity and higher PSNR can be achieved. The steps involved in the proposed algorithm are listed below –

1. Define a *qp* for starting frame.
2. Compute average MAD at n^{th} frame $MAD_{avg}(n)$ as

$$MAD_{avg}(n) = (1/n-1) \sum_{i=1}^{n-1} MAD(i) \quad \textbf{(3.4)}$$

3. Compute MAD_{cb} for this macro-block using equation 3.3.
4. A MB is said to be complex picture if it contains very detailed picture information. Complexity of a MB is quantitatively defined using a threshold-based approach.
   - Threshold for n^{th} frame T (n) is defined as 80% of $MAD_{avg}(n)$. This 80% threshold selection is done based on 20 different test sequences of different resolutions experimentally using classical decision theory [16].
   - If the MAD_{cb} for a particular MB is above this threshold, it is said to be complex. This threshold is dynamically computed and it is updated at frame layer.
   - If in a frame, $MAD_{cb} > T(n)$, for at least one MB in a frame, the frame is declared as complex and MB is chosen as the basic unit for that frame.
   - If in a frame, $MAD_{cb} <= T(n)$, for all MBs, then basic unit is chosen as frame.

3.3.3 Adaptive packetization of the encoded data

VIDEO DATA

The H.264 encoded video frames are broken into fragments of $N = (fr+H) = 1440$ *bytes* (optimal size arrived through experimentation in real network scenarios), which is transmitted at an interval of δt seconds.

$$\delta t = (fr + H) / (1000 * BW_{eff}) \quad \textbf{(3.5)}$$

where, fr is the fragment size, H is the fragment header size and BW_{eff} is the effective bandwidth computed as per Table 3.1. A 9-byte header is added to each video fragment. Details of the header for a video fragment are given below:

Frame type (1 byte) – This is the indication of the type of frame whether it is I (Independent) or P (Predictive) frame.

Total sub-sequence number (1 byte) – This is the number indicating the total number of fragments generated from a video frame.

Sub sequence number (1 byte) – This is the fragment number.

Sequence number (4 bytes) – This is the video frame number.

Video payload size (2 bytes) – This contains the number of video bytes in the current fragment.

At the transmitter end each encoded video frame is broken down to fragments of fixed size. Each of these fragments is given a sequence numbers starting from zero. This sequence number is used at the receiver side for the re-assembly of the frame from the fragments. The fragments are also given a frame identification number, which indicates to which frame it belongs.

Receiver receives and buffers the individual video fragments. Buffering is required for re-assembly of video fragments that arrive out of sequence. When all the fragments of a video frame are received, they are reordered to regenerate the original video frame using the sequence number of the fragments. To determine if all the fragments have been received or not, the information in the header about the number of fragments in the current frame is used.

The application decides to drop a frame only if a fragment belonging to a different and newer frame arrives at the receiver before all the fragments of the current frame has been received completely. The assumption here is that fragments from new frame start arriving before the completion of the current frame only if fragments belonging to the current frames are lost in transit. This method has the advantage that packets will not be discarded based on their transit delay with a frame level control - this is one of the reasons for choosing the proposed method over RTP. Thus the probability of receiving good packets on a slow network is increased.

AUDIO DATA

The 20 msec audio frames are encoded using NB-AMR audio encoder. M audio frames are aggregated and sent in a single UDP fragment. M is chosen to match the video frame rate. The DTX (Discontinuous Transmission) is enabled in the NB-AMR encoder, which indicates the silence period in the audio. If the silence period for more than D seconds then the audio transmission is discontinued.

In the proposed implementation, M is taken as 10 (matching a video frame rate of 5 fps) and D is kept configurable based on the channel condition. For a good channel (BW_{eff} > 4 kbps), D is taken to be large (more 10 seconds), for a bad channel ((BW_{eff} <= 4 kbps), D is kept small (3 to 5 seconds).

3.4 Results

3.4.1 Experimental Setup

The proposed enhancements were implemented on top of the video chat application on HIP (as described in Appendix A.2) and were tested using three different pair of network scenarios:

1. 2G Modem on both sides of video chat,
2. One side 2G Modem and 1 one side ADSL broadband,
3. Both side ADSL broadband.

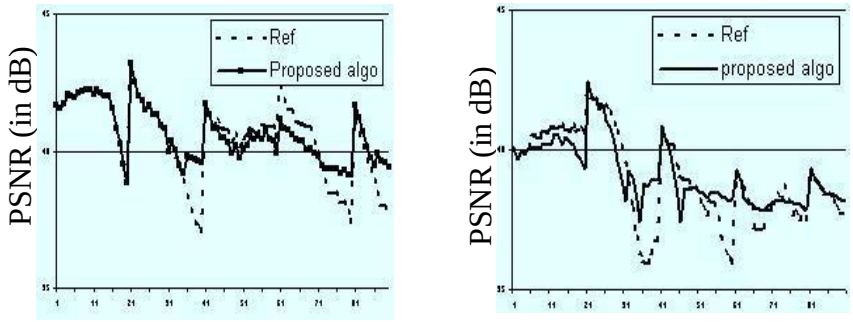
The three kinds of network scenario are chosen to address the different network conditions that can be encountered in real deployment.

The network sensing concept is tested through measuring the actual packet probe-pair delay and correlating it to the computed effective bandwidth. The video encoding efficiency is measured using standard video test sequences for video chat scenarios like Akiyo and Grandma and calculating the PSNR for both standard H.264 implementation and proposed algorithm. The overall system including the adaptive packetization is tested through judging the overall user experience for the video chat system on HIP.

3.4.2 Experimental Results

Fig. 3.3 depicts the video encoding efficiency – it presents the comparison of PSNR between the proposed video rate control algorithm and the H.264 reference code for two different video sequences (Akiyo and Grandma, both of which are representative of video chat kind of scenarios). The x-axis denotes the frame number. The results can be explained in the following way. The frame level rate control of existing H.264 reference algorithm gives better result in some high PSNR areas and worse result in complex regions. Moreover bit fluctuation in reference implementation is also very high resulting in highly fluctuating PSNRs (and hence video quality) for CBR. This fluctuation is

controlled through our proposed methodology of adaptively switching between MB and Frame as basic unit based on scene complexity and thereby preserving the bit-budget. It also gives performance benefit in terms of computational complexity and thus overall betterment in performance is achieved. After confirming the efficacy of the proposed algorithm through test sequences, the proposed algorithm was added into the H.264 encode/decode part of the HIP video chat application and was used for subsequent testing in real scenarios.



Akiyo @ 25kbps Grandma @ 30kbps

Fig. 3.3: Quality Comparison - proposed algorithm vs. standard implementation

Table 3.2 shows the effective bandwidth (kbps) in different types of network combinations and the standard deviation of the effective bandwidth measured using the proposed method as per Table 3.1. The same results are graphically represented in Fig. 3.4. It is quite obvious form the figures that Modem-Modem scenario depicts the worst possible network condition where not only the bandwidth is low, but there is also widespread fluctuation. In ADSL-ADSL scenario, the bandwidth available is large and the relative fluctuation is low – hence it is clearly the best possible network condition. The Model-ADSL scenario falls in between the two – it has low bandwidth (modem characteristics) but also low fluctuation. The Modem-Modem and Modem-ADSL scenarios are taken to further illustrate how our proposed channel condition estimation correlates with the actual channel condition.

Table 3.2: Network Statistics in Different Scenarios

Network Type	Mean	stdev
ADSL-ADSL	596.14	203.45
Modem- Modem	26.96	19.23
Modem-ADSL	18.13	3.21

In the proposed network condition sensing method, effective bandwidth is estimated based on the time difference of RTT for the probe packets. The variation in the time difference (msec) of RTT for the probe packets and the variation of estimated bandwidth (kbps) in a Modem-ADSL network and Modem-Modem network are shown in Fig. 3.5 and Fig. 3.6 respectively. The x-axis denotes the probe packet number. The correlation between estimated effective Bandwidth (which depicts the actual channel condition) and the measured RTT Delay is quite obvious in the figures, which in turn proves the efficacy of our proposed network condition estimation algorithm.

Finally, to test the overall system performance and specifically the adaptive fragmentation part, a set of 20 end users were asked to comment about their overall experience in the video chat in real network conditions under two scenarios – one with standard video chat implementation using standard audio/video codecs and standard transport protocols like RTP and another with the enhanced implementation with proposed improvements. All the users reported better audio quality and better perceived video quality for the proposed implementation. The better audio quality can be seen as an effect of the proposed fragmentation algorithm, where audio is given higher priority and video is encoded and fragmented according to rest of the available bit-budget as per effective bandwidth estimate. Perceived video quality improvement can be attributed to two factors – a) adaptive rate control scheme as proposed, and b) the video fragmentation scheme as proposed, which increases probability of receiving all packets of a frame under bad network conditions and drops a frame all-together if all packets of a video frame is not received. Since significant improvement was reported

from 100% of the users, and results from standard test video sequences also support the proposed enhancements, no separate statistical end user study results were conducted.

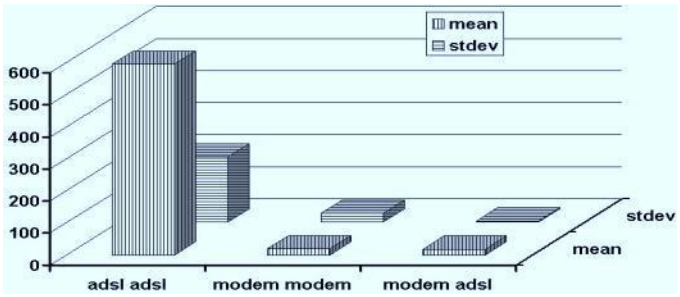


Fig. 3.4: Network Statistics in Different Scenarios

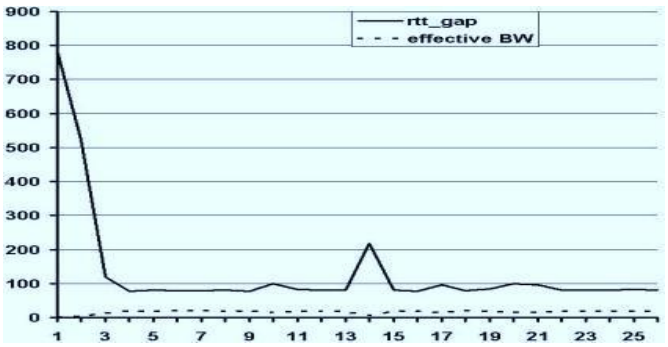


Fig. 3.5: Effective Bandwidth in Modem-ADSL network

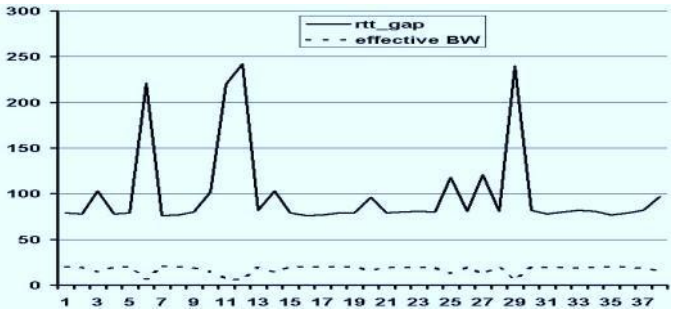


Fig. 3.6: Effective Bandwidth in Modem-Modem network

3.4.3 Discussion

In this section three novel techniques were introduced for improving experience of video chat over low-QoS wireless networks –

1. **Sensing the network condition -** An experimental heuristics based mapping of Effective Bandwidth to probe-packet-pair delay leading to a low-complexity implementation for network condition sensing.
2. **Rate control of video codec –** Adaptive selection of frame or macro-block as the basic unit for computing the quantization parameter leading to a low complexity yet acceptable performance adaptive rate distortion algorithm that is integrated with the network condition sensing system.
3. **Adaptive Packetization –** Experiment heuristics based determination of optimal packet fragment size and adaptive calculation of the inter-packet-fragment delay for both video and audio based on network condition.

The complete end-to-end system was implemented on HIP and tested on real network involving 2G modems and ADSL broadband. Results support the efficacy of heuristics based effective bandwidth calculation, shows improved video quality in spite of reduced computation complexity and improved overall end-user experience due to the adaptive packetization.

3.5 Conclusions

A multi stage adaptive rate control over heterogeneous networks for a H.264 based video chat solution is proposed in this chapter. Three stages of the system are presented – probe-packet-pair delay based sensing of network condition, adaptive rate control of audio and video compression and adaptive packet fragmentation for video and audio packets. A novel idea for each of the stages is proposed in form of an experimental heuristics based mapping of effective bandwidth to probe packet delay, a low complexity video rate control through automatic switching between frame and macro-blocks and an adaptive scheme for audio/video packet fragmentation. The system is tested in real network scenarios of 2G modems and ADSL. Results show definite improvement over the existing system.

References

[1]. Han, H. S. Park, Y. W. Choi, and K. R. Park, "Four-way Video Conference in Home Server for Digital Home", *Proc of 10th International Symposium on Consumer Electronics (ISCE '06),* Page(s) 1256-1260, Russia, 2006.

[2]. Mir Md. Jahangir Kabir, Md. Shaifullah Zubayer, Zinat Sayeeda, "Real-time video chatting in low bandwidth by Facial Action Coding System", *Proceedings of 14th International Conference on Computer and Information Technology (ICCIT 2011)*, Dhaka, Bangladesh, 2011.

[3]. Torgeir Haukaas, "Rate Adaptive Video Streaming over Wireless Networks", *Master of Science Thesis,* Norwegian University of Science and Technology, Department of Telematics.

[4]. Jacob Strauss, Dina Katabi, Frans Kaashoek, " A Measurement Study of Available Bandwidth Estimation Tools", *IMC 03,* Florida, USA, 2003.

[5]. Karthik Lakshminarayanan, Venkata N. Padmanabhan, Jitendra Padhye, "Bandwidth Estimation in Broadband Access Networks", *IMC 04,* Sicily, Italy, 2004.

[6]. J.-C. Bolot, "End-to-end packet delay and loss behavior in the Internet", *Proc. ACM SIGCOMM Symp. Communications Architectures Protocols,* Sept. 1993, Page(s) 289—298, San Francisco, CA, Sept. 1993.

[7]. S. Mascolo, "End-to-end bandwidth estimation for congestion control in packet switching networks", *US Patent No US7130268B2,* US, 2006.

[8]. S. Keshav, "Methods and apparatus for achieving and maintaining optimum transmission rates and preventing data loss in a processing system network", *US Patent No US005627970A by Lucent Technologies Inc,* US, 1997.

[9]. J. Matta, and R. Jain, "Method and apparatus for quality of service determination", *US Patent No. US20100177643 A1,* US, 2010.

[10]. Proposed Draft of Adaptive Rate Control, *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG,* JVT-H017, 2002.

[11]. G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression", *IEEE Signal Process. Mag.,* November 1998.

[12]. K. Yu, J. Li, and S. Li, "Bi-level and full-color video combination for video communication", *US Patent No. US007912090B2 by Microsoft Corporation,* US, 2008.

[13]. 3GPP TS 26.071 version 10.0.0 Release 10, ETSI TS 126 071 V10.0.0, "Mandatory speech CODEC speech processing functions, AMR speech Codec", 2011-04.

[14]. ITU-T Rec. H.264, "Advanced Video Coding for Generic Audiovisual Services", 2010.

4
Low-complexity Video Security

4.1 Introduction

In last a few years, internet technology and video-coding technology has grown very significantly. With this trend of internet based multimedia applications, digital rights management and security has become essential for copyright management and access control. With relation to HIP, the applications like Remote Medical Consultation and Distance Education has sensitive and rights-protected multimedia data (**Chapter 2** and **Appendix A**). For example, the video chat session between doctor and patient contains sensitive data that needs to be protected through access control. Similarly, the multimedia content for distance education needs both access control and digital rights management.

Encryption is the standard way for providing access control to the multimedia content like a video chat session or an education tutorial video, while watermarking is the standard way to provide digital rights management (DRM) for copyright protected content like education tutorials. While there are many established encryption and watermarking techniques available, they introduce significant computation overhead especially for real-time systems running on a low processing power platform like HIP. Hence there is need for low-computational-complexity video encryption and watermarking algorithms without compromising on the security.

In section 4.2, a novel low-complexity watermarking algorithm is introduced as a DRM tool that can be embedded inside the video compression algorithm itself. In section 4.3, a novel low-complexity encryption algorithm is proposed that is embeddable inside the video compression algorithm. In each of sections 4.2 and 4.3, the problem definition is elaborated first through literature study, followed by proposed algorithm description and then results with discussion are presented. Finally in section 4.4 a conclusion and summary is provided.

4.2 Low-Complexity Video Watermarking

4.2.1 Problem Definition

Watermarking is the process that embeds information (called watermark) into a multimedia object with or without visual artifacts, making the information an integral part of the multimedia object. From DRM perspective, typically a watermark can be information like

- A serial number or random number sequence
- Ownership identifiers

- Copyright messages
- Transaction dates
- Logos in form of binary or gray level images
- Text strings

For the distance education application on HIP, the IP address / identity of the content creator and consumer, date/time stamp and a logo image can be watermarked inside the video content to provide DRM and the watermark needs to be invisible. A good invisible watermarking technique has the following requirements –

- Should convey as much information as possible.
- Should be secret and accessible to authorized parties only.
- Should withstand any signal processing and hostile attacks, i.e. should be robust
- Should be imperceptible.

It is interesting to note here that Robustness and Imperceptibility are contradicting requirements and hence poses a major challenge in watermarking technique design. In addition, there is need for such watermarking systems to be low complexity, without which it is difficult to put them in low power real-time embedded systems like HIP.

It should be noted here that since both watermark embedding and extraction is in our control in applications like distance education (teacher's workbench and student station, as described in Appendix A.2), one can afford to propose to use proprietary schemes as long as they show computational efficiency and withstand relevant attacks. In scenarios where both sides are not in our control (like IPTV streaming), it is imperative to use standard-based systems and hence the proposed solution cannot be applied.

H.264/AVC is used as the video compression technology as it is the most efficient compression scheme ([1], [2]). In [3], a very good overview of video watermarking techniques and its applications is given. In his famous paper in [4], Hartung provides excellent technology overview of video watermarking techniques. However both [3] and [4] do not deal with H.264, a relatively new video compression standard, as they were written before H.264 came up. The same can be said about [5], which though dealing with compressed domain watermarking, supports only MPEG2 compression standard. Reference [6] proposes a hybrid watermarking scheme that provides both fragileness and robustness for H.264 compressed videos at the cost of some reduced efficiency and higher computational complexity. However, for the given education DRM application, the concern is more about efficient implementation of robust watermarks, hence the scheme proposed in [6] is not suitable. In [7] and [8], a low-complexity implementation of watermarking in H.264/AVC is given, but they do not provide analysis for neither perceptual video quality ([9]), nor do they provide detailed attack analysis.

The main contribution of the work presented in this section is to provide a robust, low-computational-complexity compressed-domain video watermarking system that can be very easily embedded inside a H.264/AVC codec without deteriorating the perceptual video quality. The complete watermarking system developed in this work not only covers the main watermark embedding algorithm, but also includes the algorithms necessary to cater to the overall system requirements like integrity check, finding optimal location inside the video for embedded the watermark etc. A modified version of H.264 Intra-prediction mode calculation is also proposed to take care of watermarking requirements.

For evaluating the security robustness of the proposed algorithm, a watermark attack evaluation tool has also been developed. As a further contribution, a novel methodology for evaluating the watermark performance against different attacks using the tool is presented. A couple of papers have already being published on the work done (Appendix B - [3], [12]). A patent is also filed on the watermark evaluation methodology (Appendix B – [c]), which is already granted in USA.

4.2.2 Proposed Watermarking Algorithm

Fig. 4.1 describes the standard process of H.264 based compression and decompression [1], [2]. The dark colored blocks in the Figure represent the additional blocks proposed here for embedding watermark within the standard H.264 video encoder.

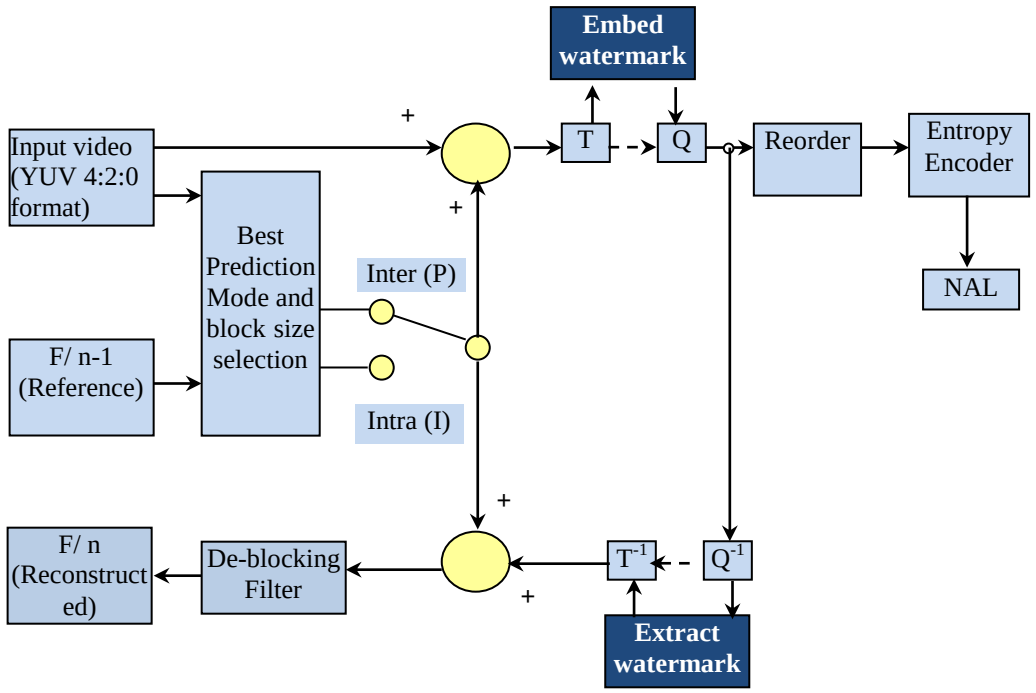


Fig. 4.1: H.264 Encoder Architecture for Watermark Embedding

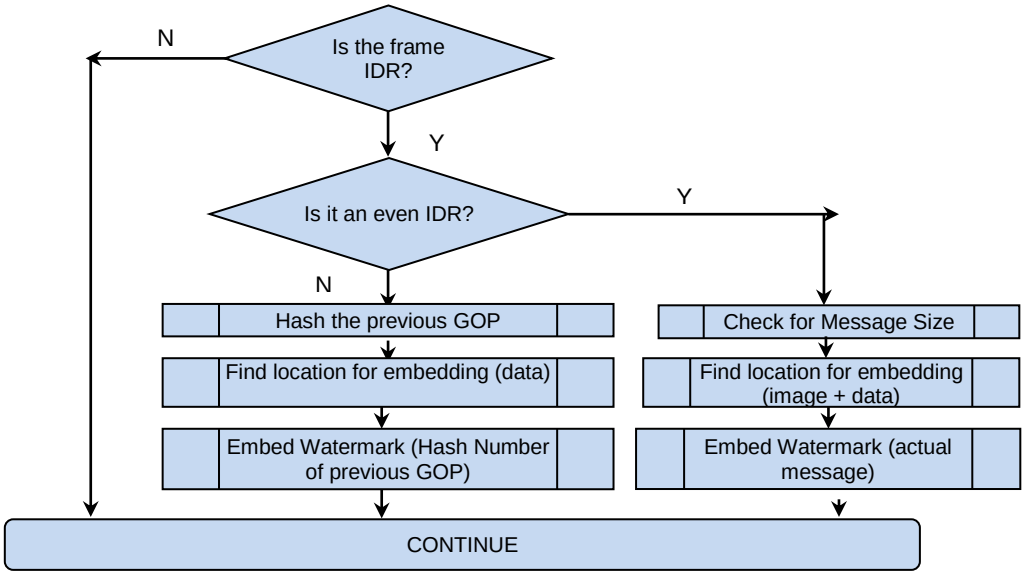


Fig. 4.2: Watermark Embedding Algorithm Flow Diagram

The overall methodology for watermark embedding is introduced in Fig. 4.2. In even numbered Independent Decoder Refresh (IDR) frames, the desired message is embedded; and in odd numbered frames the watermarking bit-stream is obtained by hashing the last Group of Picture (GOP) to ensure integrity. This scheme is required to improve robustness of the system against attacks. The image to be watermarked is normally a small logo in binary form. The data to be watermarked is typically identifiers like number, IP address, and text string etc. all represented as byte arrays. The proposed methodology is elaborated in detail below.

CHECK FOR MESSAGE SIZE

Typically the logo image watermark has significantly more bits than the text watermark and for sake of robustness the image needs to be embedded inside the diagonal macro-blocks (details in next section – Find Location for Embedding). Hence, check for message size boils down to checking whether there are enough no. of diagonal macro-blocks to embed the logo image (as proposed in the next section, each macro-block can hold one bit of image watermark data). This process of checking the size is depicted in Fig. 4.3.

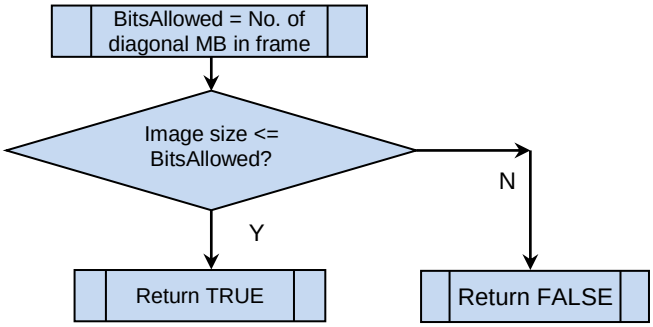


Fig. 4.3: Flow Diagram for Checking the Watermarking Message Size

FIND LOCATION FOR EMBEDDING

The watermark is embedded in the compressed domain by altering some of the DCT coefficients in the macro-blocks. Finding location amounts to finding which DCT coefficient in which macro-block is to be selected. The selection is based on the following observations:

1. DCT coefficients are zero in most cases
2. Most significant information lies in top and left
3. Modification of diagonal elements at right and bottom results in insignificant artifacts
4. Coefficients in diagonal positions are more stable than the others.

Since the logo image has more information to embed and is sensitive to change, for the sake of robustness and imperceptibility, it is imperative to embed the image in the diagonal macro-blocks (observation 3 and 4). In line with observation 2, the 10^{th} and the 15^{th} DCT coefficient can be chosen as the candidate sub-block (SB) location within the diagonal macro-blocks. Similarly, the data part of the watermark is embedded in 10^{th} and the 15^{th} DCT coefficient of the ab-diagonal macro-blocks. The candidate locations are depicted in Fig. 4.4.

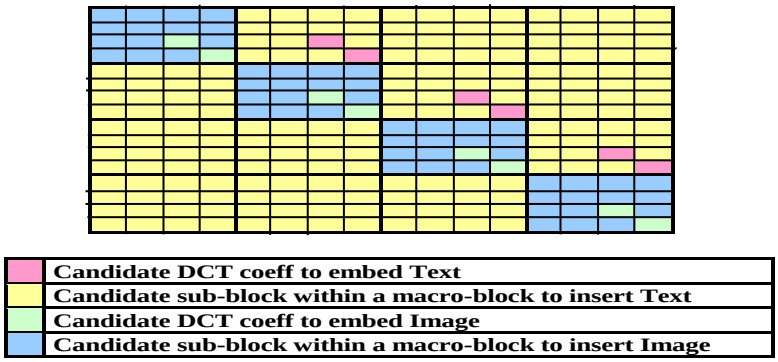


Fig. 4.4: Location Selection for Embedding inside Image

In both cases of image and data, the watermark bit-stream is embedded using the following principle –

If the index of the watermark bit to be embedded is odd,

Embed it in 10^{th} coefficient

Else

Embed it in 15^{th} Coefficient

The overall methodology is introduced in Fig. 4.5.

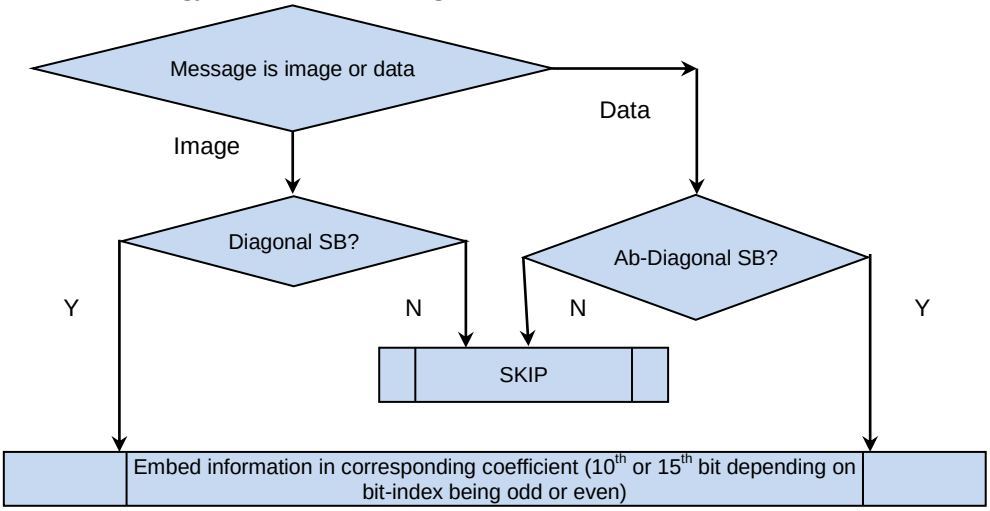


Fig. 4.5: Algorithm Flow for Handling Data or Image for Watermarking

EMBED WATERMARK

As a typical example, a *HxW* logo image in binary format is used for watermarking along with a *K* byte data information. Hence the total number of bits to embed comes out to be *N = HxW + K*8.* This information is stored in an *N* byte array (called w_n) whose each byte is 0/1. Then w_n is quantized using same quantization parameter (*qp*) as used in the H.264 video compression and the quantized values are stored in another array (w_{qn} of size *N*). For each w_{qn}, the location of embedding inside the image (as already proposed) is found. The image location mapped w_{qn} is depicted as *M(u,v),* where *(u,v)* denotes the position in the DCT domain for a given frame. Watermark bits are inserted by altering the quantized AC coefficients of luminance blocks within I-frames. In order to survive the transcoding attacks, the watermark signal *M(u,v)* must be strong enough to survive the quantization, so that

$$|M_q(u, v)| = |quant[M(u, v), qp]| \quad (4.1)$$

where the quant[·] denotes the quantization operation, *qp* denotes the quantization parameter (typically 0 to 51 for H.264). Obviously, greater the *M(u,v),* higher is chance of the watermark to survive the requantization during transcoding. It should be noted here that the imperceptibility requirement is already addressed while choosing the watermarking location and now focus is given on the robustness aspect.

Representing the original value of the quantized DCT coefficient of the luminance component of the I frame as $X_q(u,v)$, the following is proposed as the watermark embedding mechanism - Replace $X_q(u,v)$ by the watermarked coefficient *Xq*, where

$$X_q = \begin{cases} max\{X_q(u, v), M_q(u, v)\} \text{ if } w_n = 1 \\ 0 \text{ if } w_n = 0 \end{cases} \quad (4.2)$$

It should be noted that $X_q(u, v)$ is cleared if '0' is embedded. It can be justified by the fact that the $X_q(u, v)$ is zero in most cases, especially in the locations chosen. Hence, it will not introduce significant artefacts.

Two basic intra-prediction modes INTRA-4x4 and INTRA-16x16 are used in H.264/AVC, comprising 4 x4 and 16 x16 block-wise directional spatial predictions, in which the pixel value of the current block is predicted by the edge pixels of the adjacent blocks. Then, the prediction error is transformed primarily by a new 4x4 integer DCT instead of the float 8x8 DCT, which is widely used in existing standards. While the smaller block-size is justified by the improvement of prediction capabilities by using the above mentioned prediction modes, it makes the embedded

watermark more sensitive to attacks or transcoding. Hence there is a need to choose an optimal mode for intra-prediction of the modified frame keeping the distortion and bit budget in mind.

The best intra-prediction mode for a watermarked macro-block S_k^* is selected by minimizing the modified Lagrange optimization function:

$$J_{Mode} = D_{REC}(S_K^*, I_k) + \lambda_{MODE} R_{REC}(S_k^*, I_k) \quad (4.3)$$

where, D_{REC} and R_{REC} represent the distortion and the number of bits, respectively, encoded for modes $I_K \in \{INTRA\text{-}4\times4, INTRA\text{-}16\times16\}$ and λ_{MODE} is Lagrange parameter.

4.2.3 Results

The complete system was implemented on the HIP platform using the 64x DSP core of the Texas Instruments' DM6446 Davinci chipset. A 24x16 logo image, date and timestamp string and IP address were taken as objects for watermarking. From the perspective of the framework introduced in **Chapter 2** and referring to Fig.2.2, the watermarking embedding and extraction will be part of the processing sub-system. The results obtained from the system are presented with respect to computational complexity, perceptual quality loss for video and security analysis.

COMPUTATIONAL COMPLEXITY

Table 4.1 and Table 4.2 give the theoretical (for embedding only) and measured computational complexity (for both embedding and extraction) of the proposed algorithm. It is obvious that watermark extraction will have less complexity compared to watermark embedding due to less overhead in message size check, actual watermark retrieval and mode prediction. Both theoretical and measured results indicate towards the low computational complexity of the algorithm. It should be noted here that in the distance education scenario, watermark embedding will be done at the Teacher's workbench, which is not a constrained platform. Only watermark extraction is done on HIP, which has even lesser, almost insignificant additional computational overhead compared to original H.264.

Table 4.1: Theoretical Computational Complexity

Operation	No. of Operations per GOP (1 GOP = 15 frames = 1 second for 15 fps video)
ADD	2779
MULTIPLY	3564
DIVIDE	1980
MODULO	3564
CONDIONAL	7524
MEMORY I/O	1584

Table 4.2 Measured Complexity

Function	CPU Mega Cycles taken per GOP (1 GOP = 15 frames = 1 second for 15 fps video)
Watermark Embedding	6.8
Watermark Extraction	3.8

QUALITY LOSS AFTER WATERMARKING

An exhaustive study of perceptual quality measurement is available in [9]. As per the results shown in [9], it is clear that Peak Signal to Noise Ratio (PSNR) can be taken as a representative perceptual quality measure. Fig. 4.6 depicts the loss of quality in the original video after watermarking in terms of (PSNR). As seen from the results, there is no loss in U and V components (which is expected as the chroma components are not disturbed) and loss in Y component is also negligible (less than 1 dB).

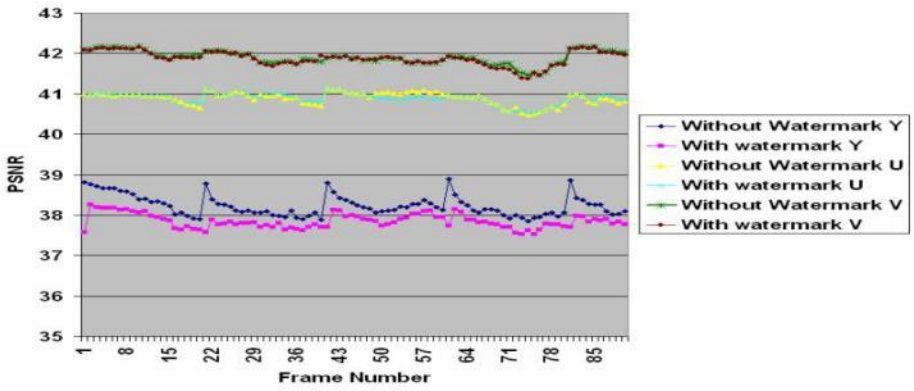


Fig. 4.6: Quality Degradation after Watermarking

SECURITY ANALYSIS

The proposed watermarking was subjected to popular watermarking attacks like Averaging attack (AA), Circular averaging attack (CAA), Rotate attack (RoA), Resize attack (RsA), Frequency filtering attack (FFA), Non-linear filtering attack (NLFA), Gaussian attack (GA), Gama correction attack (GCA), Histogram equalization attack (HEA) and Laplacian attack (LEA)[9].

The watermarked video is evaluated against all the different attacks by using a Watermarking Evaluation Tool. The architecture of the tool is given in Fig. 4.7. It should be noted that even though the tool is used to evaluate our H.264 based watermarking scheme, it is generic enough to support any video watermarking system.

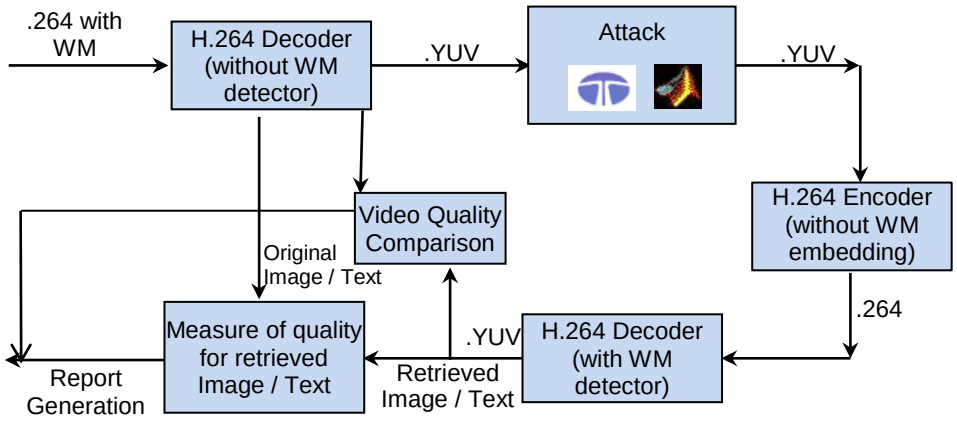


Fig. 4.7: Architecture of Watermarking Evaluation Tool

Based on the above observations, three metrics for judging the goodness of watermarking scheme are considered – quality of the video after attack (perceptual quality measure), retrieved image quality after attack and retrieved text quality after attack (robustness measures). Each of these methodologies is presented in detail below. Finally, a novel aggregated decision logic using the above three metrics are proposed for evaluating the overall performance of the watermarking system against attacks.

Quality of Video after Attack

There are quite a few objective video quality measures available[10]. A total of 10 different measures are taken here to judge the video quality –

1. Average Absolute Difference (AAD)

[9] http://cvml.unige.ch/publications/postscript/99/VoloshynovskiyPereiraPun_eww99.pdf
[10] http://atc.umh.es/gatcom/bin/oqam/Referencias/Wang_Sheikh_Bovik_BookChapter_2003.pdf

2. Mean Square Error (MSE)
3. Normalised Mean Square Error (NMSE)
4. Laplacian Mean Square Error (LMSE)
5. Signal to Noise Ratio (SNR)
6. Peak Signal to Noise Ratio (PSNR)
7. Image Fidelity (IF)
8. Structural Content (SC)
9. Global Sigma Signal to Noise Ratio (GSSNR)
10. Histogram Similarity (HS)

There are other measures like Maximum Difference, Norm, Average Absolute Difference, L–Norm, Normalised Cross-Correlation, Correlation Quality, Sigma Signal to Noise Ratio, Sigma to Error Ratio etc. which are typically used as metrics for distortion in image domain, but it was found that they do not convey much information as far as video quality is concerned. In our proposed methodology, each of the frames of the attacked video sequence is evaluated against each of watermarked video sequence. Each frame is judged by using the 10 different measures stated above, which are then combined using weighted score as outlined below –

1. First calculate the above ten metrics for three pairs of videos – a) Two identical videos, b) Two completely different videos and c) original video and same video going through a compression decompression chain. These three pairs represent two extreme bounds and the average case as far as video quality difference is concerned. It is observed that five metrics (*AAD, GSSNR, LMSE, MSE, PSNR*) vary largely for three types of cases. But other metrics do not vary that much. So more weightage can be given to these five metric to arrive at a combined metric

   $W\_VAL = ((AAD+GSSNR+LMSE+MSE+PSNR)*3+HS+IF+NMSE+SC+SNR)$
2. 14 test video streams were taken and subjected to different kinds to attacks and the evaluation tool was used to calculate *W_VAL* for each scenario. Parallely, 20 users were requested to judge attacked and original watermarked video sequence based on their perception. This judgement is purely based on human vision psychology (HVS). All these opinions are summed up in Mean Opinion Score (MOS). Finally, based on MOS, a fuzzy value is assigned to the parameter C_{qual} that is calculated from *W_VAL*.

$$\begin{aligned} &IF\ (W\_VAL >= 90), && C_{qual} = Excellent \\ &ELSEIF\ (W\_VAL >= 80), && C_{qual} = Good \\ &ELSEIF\ (W\_VAL >= 75), && C_{qual} = Average \\ &ELSEIF\ (W\_VAL >= 70), && C_{qual} = Bad \\ &ELSE && C_{qual} = Poor \end{aligned} \quad \textbf{(4.4)}$$

The Table 4.3 shows the results of the evaluation for the proposed Watermarking algorithm using the methodology presented above along with image snapshots of one of the test sequence (“Stephan”) in Fig. 4.8.

Table 4.3: Video Quality after Attacks

Attack	W_VAL	C_{qual}
AA	100	Excellent
CAA	52	Poor
FFA	25	Poor
GCA	27	Poor
GA	71	Bad
HEA	27	Poor
LEA	28	Poor
NLFA	25	Poor
RsA	100	Excellent
RoA	37	Poor

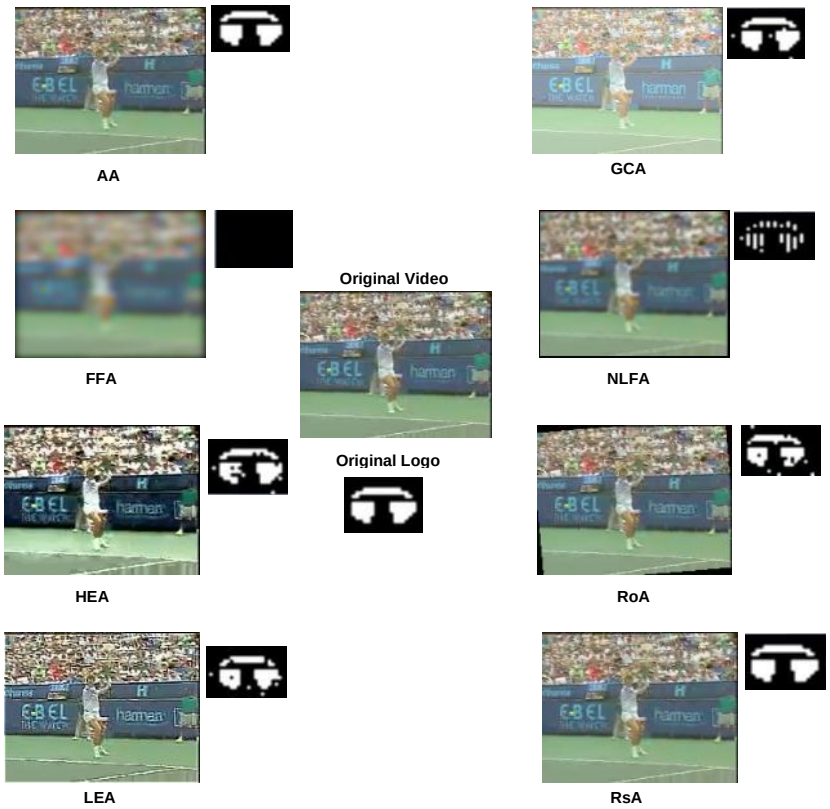


Fig. 4.8: Snapshots of Video Frames and Retrieved Logo from Watermarked Video

Quality of retrieved image

After extracting the watermarked image, retrieval quality is judged by using the parameters like bit error, deviation of Centroid and difference in crossing count. A combined metric calculation based on MOS is proposed as below -

1. Compute Euclidian distance (d) of Centroid of black pixels of retrieved and original binary image and normalize it with height (h) and width (w) of the image to arrive at the deviation parameter (d_e) -

$$\boldsymbol{d_e} = \frac{\boldsymbol{d}}{\sqrt{\boldsymbol{h}^2 + \boldsymbol{w}^2}} * 100$$

2. Bit error (b_e) is the number of bits differing between retrieved and original binary image represented in percentage.
3. If c be the difference in crossing count of 0 to 1 of original and retrieved binary image, Crossing count error (c_e) is defined as:

$$\boldsymbol{c_e} = \frac{\boldsymbol{c}}{\boldsymbol{h} * \boldsymbol{w}} * 100$$

4. Final error in retrieved image is defined as:
   $\boldsymbol{e} = (\boldsymbol{c_e} + \boldsymbol{b_e} + \boldsymbol{d_e})/3$
5. Based on MOS introduced in the previous section, the following decision logic can be arrived at for the quality of the retrieved image (C_{img})

   IF e < 0.5 *Cimg = Excellent*
   IF 5 > e > 0.5 *Cimg = Good*
   IF 10 > e > 5 *Cimg = Medium*
   IF 15 > e > 10 *Cimg = Bad*
   ELSE *Cimg = Poor*

(4.5)

Table 4.4 gives the results of the retrieved image quality after attacks using the tool.

Table 4.4: Retrived Image Quality after Attacks

Attack	b_e	c_e	d_e	e	Image Quality (C_{Img})
AA	0.000	0.000	0.000	0.000	Excellent
CAA	5.469	9.896	3.448	6.271	Medium
FFA	5.469	10.938	55.172	23.860	Poor
GCA	0.781	1.563	3.448	1.931	Good
GA	4.948	9.896	24.138	12.994	Bad
HEA	1.563	1.563	3.448	2.191	Good
LEA	1.823	2.083	0.000	1.302	Good
NLFA	5.729	10.417	13.793	9.980	Medium
RsA	0.000	0.000	0.000	0.000	Excellent
RoA	0.781	0.521	0.000	0.434	Excellent

Quality of the retrieved text

After extracting the watermarked text, retrieval quality is judged by using popular distance measures for string comparison like Hamming distance[11] and Levensthein distance[12]. The average of the above two scores and the MOS can be used to arrive at a single metric as outlined below.

1. Compute the Hamming distance ($\boldsymbol{h}$) and Levensthein distance ($\boldsymbol{l}$).
2. The mean error ($\boldsymbol{t_e}$) is computed as:

$$t_e = \frac{l + h}{2}$$

3. In the same way as previous two cases, the MOS based retrieved text quality C_{txt} is as follows -

$$\begin{aligned} &IF\ t_e < 0.5 && C_{txt} = Excellent \\ &IF\ 1 > t_e > 0.5 && C_{txt} = Good \\ &IF\ 3 > t_e > 1 && C_{txt} = Medium \\ &IF\ 5 > t_e > 3 && C_{txt} = Bad \\ &ELSE && C_{txt} = Poor \end{aligned} \qquad \textbf{(4.6)}$$

Table 4.5 gives the results of the retrieved text quality after attacks using the evaluation tool.

Table 4.5: Retrieved Text Quality after Attacks

Attack	L	H	t_e	Text Quality (C_{txt})
AA	0	0	0.000	Excellent
CAA	6	1	3.5	Bad
FFA	6	1	3.5	Bad
GCA	0	1	.5	Good
GA	5	1	3	Bad
HEA	6	7	6.5	Poor
LEA	4	5	4.5	Bad
NLFA	6	1	3.5	Bad
RsA	0	0	0.000	Excellent
RoA	0	0	0.000	Excellent

[11] http://www.lee.eng.uerj.br/~gil/redesII/hamming.pdf
[12] http://profs.sci.univr.it/~liptak/ALBioinfo/files/levenshtein66.pdf

Aggregated Decision Logic

In descriptive terms, the aggregation can be performed using the concept that if there is no significant degradation in video quality and the retrieved watermarked information does not contain significant errors, then the watermarking scheme has a high measure of goodness. An aggregation scheme is proposed in Table 4.6 with results from equations (4.4), (4.5) and (4.6).

Table 4.6: Overall decision making process and Guideline

Video Quality	Retrieved Image Quality	Retrieved Text Quality	Overall Measure of Goodness
Excellent or Good	Excellent	Excellent	Excellent
Excellent or Good	Excellent	Good	Good
Excellent or Good	Good	Excellent	Good
Excellent or Good	Good	Good	Good
Excellent or Good	Medium	Medium	Medium
Excellent or Good	Bad or Poor	Medium	Bad
Excellent or Good	Medium	Bad or Poor	Bad
Excellent or Good	Bad or Poor	Bad or Poor	Poor
Medium, Bad or Poor	Any	Any	Attack degrades video quality beyond acceptable limit – hence attack itself is not suitable and hence need not be considered

Using the results from Table 4.3, Table 4.4 and Table 4.5, the aggregation logic elaborated in Table 4.6 is applied to arrive the overall performance of the proposed algorithm against different attacks. The results are presented in Table 4.7.

4.2.4 Discussion

As seen from the results in Table 4.7, in 8 out of 10 attacks, the video quality is degraded beyond acceptance level. Hence these attacks are not of concern for content-rich applications requiring DRM, like distance education. Under the two remaining attacks, the performance of the proposed algorithm shows excellent robust behavior. It is interesting to note that out of the 8 attacks, even after degrading the video quality level beyond acceptance, 6 cases produce acceptable quality for the retrieved image and 2 cases produce acceptable quality for the retrieved text. This further reinforces the robustness of the proposed algorithm. The observed behavior also suggests that the image watermarking to be more robust than the text watermarking – this property can possibly be exploited by encoding the text in an image like QR-codes before embedding as watermark.

The results also suggest that there is negligible degradation in video quality after watermark insertion (Fig. 4.6) and the computational overhead of the watermark extraction is negligible (Table 4.2). Hence the algorithm can be comfortably used for DRM in the Distance Education solution on HIP.

Table 4.7: Result Summary for the proposed Algorithm under different attacks

Attack	Video Quality (C_{qual})	Image Quality (C_{Img})	Text Quality (C_{txt})	Overall Performance against Attack
AA	Excellent	Excellent	Excellent	Excellent
CAA	Poor	Medium	Bad	Attack Degrades Video Quality
FFA	Poor	Poor	Bad	Attack Degrades Video Quality
GCA	Poor	Good	Good	Attack Degrades Video Quality
GA	Bad	Bad	Bad	Attack Degrades Video Quality
HEA	Poor	Good	Poor	Attack Degrades Video Quality
LEA	Poor	Good	Bad	Attack Degrades Video Quality
NLFA	Poor	Medium	Bad	Attack Degrades Video Quality
RsA	Excellent	Excellent	Excellent	Excellent
RoA	Poor	Excellent	Excellent	Attack Degrades Video Quality

4.3 Low Complexity Video Encryption

4.3.1 Problem Definition

Any kind of application that needs video sharing in one form or other requires video encryption for preventing illegal access. For example, the video chat application used as remote medical consultation tool between the patient and doctor on HIP needs to be secure so that the content of the video chat is not accessible to eavesdroppers or unauthorized entities. In such scenario, one needs to employ some form of encryption at the video content level in addition to standard network level security. A similar requirement exists for preventing unauthorized access of the multimedia content for the Distance Education application. In order to be implemented in a low-cost system like HIP, the encryption system needs to be robust yet computationally less expensive.

Encryption in uncompressed domain is similar to data encryption, however it requires significant computational complexity and introduces overhead in form of decryption of the whole streamed data to have access to the video headers. Hence typically video encryption is done in compressed domain. Description of compressed domain video encryption systems can be found in [10] to [12]. In [10], one gets a good overview of compressed domain video encryption schemes. In [11] and [12], various techniques for compressed domain video encryption like multi-layer coding and motion vector estimation are discussed. But these techniques neither use H.264/AVC [1], the video compression technology used in HIP for its efficiency, nor do they focus on reducing computational complexity, both requirements being of primary importance in the context of HIP applications. In [13] and [14], the authors propose low computational complexity compressed domain video encryption schemes based on various techniques like spatial shuffling and sign bit encoding, but their work focuses on MPEG standards and not on H.264/AVC.

Works on compressed domain H.264 encryption can be found in [15] and [16]. In [15], Yuan Li et. al. uses the method of scrambling the intra-prediction, inter-prediction mode and encryption of motion vectors. Their results show that their proposed method exhibits good security properties and has low impact on compression ratio. But the problem with their method is that it needs an overhead of keeping a decision block that controls the prediction modes and motion vectors and hence is computationally intensive. In [16], Yuanzhi Zou et. al. developed their algorithm based on the analysis of H.264 entropy coding, and the algorithm has the features of being irrelative to individual ciphers and adaptive to digital right management. Their scheme has good security properties and can greatly reduce the hardware design complexity of decryption process.. However, their algorithm is designed more for storage applications and not for streaming applications required in HIP. It should be noted here that since both encryption and decryption is in our control for applications like video chat, one can propose to use proprietary schemes as long as they show computational efficiency and are sufficiently secure.

The main contribution of our work presented in this section is to provide a low-computational-complexity two-stage compressed-domain video encryption algorithm that can be easily embedded inside a H.264/AVC codec. This is achieved through keeping the header encryption separate and novel reuse of the flexible macro-block re-ordering (FMO) feature of H.264/AVC as the encryption operator. As an additional contribution, other than providing the security and complexity analysis, the effect of the encryption-decryption chain on the video quality is also presented here, which is not found in works reported in the literature, but is important from end-user experience perspective. A paper has been published on the work done (Appendix B - [4]) and a patent has been filed (Appendix B - [d]).

4.3.2 Proposed Encryption Algorithm

In the proposed encryption method, encryption of H.264 encoder is achieved in two folds - conventional encryption method for encrypting header and encryption the video content using the architecture of H.264 AVC.

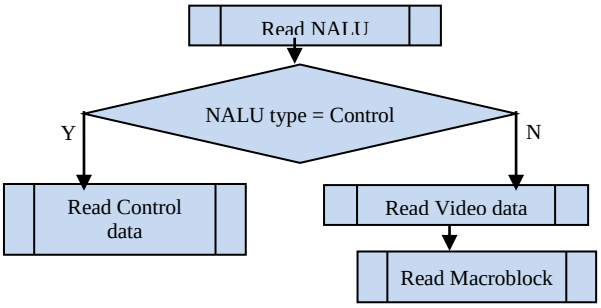


Fig. 4.9: Block diagram of Read module

Basic unit of processing of H.264 data in decoder side is known as Network Abstraction Layer Unit (NALU). Any NALU can contain either video data or control data. The syntax of the read module is shown in Fig. 4.9. Some control information like profile, label, height, and width are written in the header part. The formation of NALU depends on application and network layer, also. For example one can think of a scenario where the encoded data is transmitted over a lossy channel. There is a high chance of packet loss and thus loss of video or control data in this case. This is taken care of by sending multiple control data (like Sequence Parameter Set (SPS) and Picture Parameter Set (PPS)) and Independent Decoder Refresh (IDR) frames because if one header gets corrupted it can recover as soon as the next set of SPS and PPS arrives.

On the other hand in the asynchronous application like video chat, synchronization requires the presence of multiple SPS, PPS. Similarly, in case of video storage (which is required in the distance education application), fast forward and rewind utility requires the presence of IDR at every 1 or 2 seconds. Some typical NALU organizations are as given in Fig. 4.10 and Fig. 4.11. The NALU type can be any of the following shown in Table 4.8 as suggested by H.264 standard [1].

Fig. 4.10: NALU organization video conferencing application

Fig. 4.11: NALU organization for video storage application

Table 4.8: NALU Unit type

NALUtype	Content of NAL unit and RBSP syntax structure
0	Unspecified
1	Coded slice of a non-IDR picture, slice_layer_without_partitioning_rbsp( )
2	Coded slice data partition A, slice_data_partition_a_layer_rbsp( )
3	Coded slice data partition B, slice_data_partition_b_layer_rbsp( )
4	Coded slice data partition C, slice_data_partition_c_layer_rbsp( )
5	Coded slice of an IDR picture, slice_layer_without_partitioning_rbsp( )
6	Supplemental enhancement information (SEI), sei_rbsp( )
7	Sequence parameter set, seq_parameter_set_rbsp( )
8	Picture parameter set, pic_parameter_set_rbsp( )
9	Access unit delimiter, access_unit_delimiter_rbsp( )
10	End of sequence, end_of_seq_rbsp( )
11	End of stream, end_of_stream_rbsp( )
12	Filler data, filler_data_rbsp( )
13 .. 23	Reserved
24 .. 31	Unspecified

As one of the novel contributions, the proposed encryption method first encrypts these SPS, PPS and IDR (called Header Encryption) before applying encryption in video data (called Video Content Encryption) whose design is elaborated in detail below.

HEADER ENCRYPTION

- Take a 16-bit Key (K_U) that will be shared through secured medium
- Encode the first frame using conventional H.264 encoder
- Take the length of IDR (l_{IDR}) - It is a 16-bit number for QCIF resolution (176x144))
- Define encryption key value K_P using a Hash function (H) of l_{IDR} and K_U

$$K_P = H(K_U, l_{IDR}) \quad \textbf{(4.7)}$$

- Mask SPS, PPS, IDR with this K_P

VIDEO CONTENT ENCRYPTION

The proposed video content encryption scheme can be explained with reference to Fig. 4.1, which describes the basic H.264 encoder architecture.

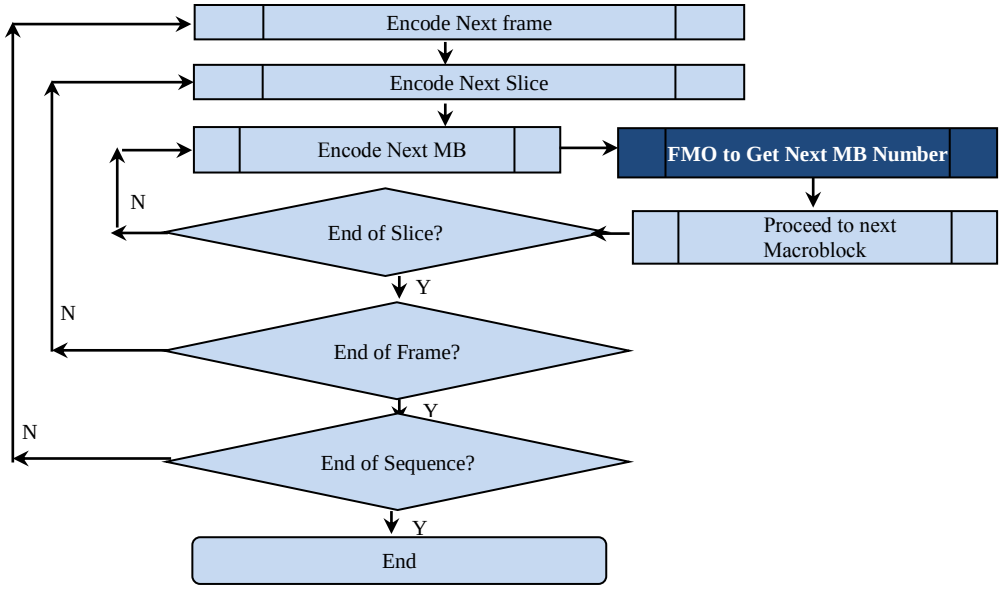


Fig. 4.12: Flow chart of the process of encoding

The proposed encryption algorithm for picture level works at the Flexible Macroblock Ordering (FMO) level (marked in dark color in Fig. 4.12 and elaborated in Fig. 4.13). For maintaining simplicity of design and reducing time complexity of the overall system, it is assumed that there is no slice partition in a frame, so that, in the existing H.264 Encoder system (without encryption), only one map unit type 0 is used. In the proposed design for implementing encryption into H.264 encoder the FMO design is modified as in Fig. 4.13 through introduction of new block for modifying FMO through key based look-up. This design is elaborated as below.

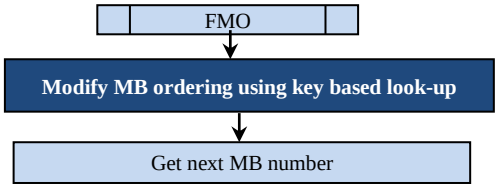


Fig. 4.13: Modified FMO algorithm

Modify macro-block ordering using key based look-up

- Start with K_P as obtained in (4.7)
- Use K_P as seed to generate random sequence L_e that has a value between 0 to 97. The last MB no. is kept as 98 as per the single slice requirement. This is used for the first GOP. For subsequent GOPs, K_P of the previous GOP is used as K_U in (4.7) to generate a new K_P.

- In encoder side MBs in an IDR frame are encoded in the order specified by a look-up table L_e. The MB to be encoded (current MB) is predicted from image samples that have already been encoded.

4.3.3 Results

The algorithm has been implemented on HIP 64x DSP co-processor. The results are presented in three aspects - security analysis, complexity analysis in form of computational and memory complexity and video quality analysis.

SECURITY ANALYSIS

In the proposed algorithm, MBs of a video frame are encoded in the order specified by a look-up table L_e which is a random permutation of the sequence {0, 1, 2, ... ,97}. But the decoder can decode the frames only if the look-up table at the decoder side (L_d) is identical to L_e. While decoding a particular MB, the decoder may not get the MB information using which the current MB was predicted if L_d is not equal to L_e .

If the hacker wants to generate the decoder look-up table L_d in brute-force, he needs to make 98! = 9.42 x 10^{153} attempts to successfully decode the video. In practical, however, this is restricted by the 32 bit key, which means L_e can be generated in 2^{32} ways, requiring half that number of attempts to decode the first GOP. Since the key is changed every GOP, the proposed encryption method is robust enough in comparison to [15] where the authors have claimed that a hacker can break the security in 2128 attempts.

COMPLEXITY ANALYSIS

Computational Complexity

Time complexity of the encryption algorithm is majorly contributed by the random number generation algorithm. The results of our proposed algorithm are presented in Table 4.9 in terms of basic operations like ADD, MULTIPLICATION, DIVISION, MODULO.

Table 4.9: Computational Complexity of the proposed algorithm

Operation	Frequency per GOP	% Increase wrt H.264/AVC
ADD	2*24*97 + 3 = 4659	0.004
MULTIPLICATION	5*24*97 = 11,640	0.200
DIVISION	5	0
MODULO	4*24*97 + 4 = 9316	0

It is seen from the results that the computational overhead due to encryption is negligible

Table 4.10: Memory complexity of the proposed algorithm

Resolution (wxh)	Picture size in MBs (wxh/(16x16))	Additional Memory requirement (in bytes)
QCIF (176x144)	99	198
CIF (352x288)	396	792
VGA (640x480)	1200	2400
SDTV-525 (720x480)	1350	2700
SDTV-625 (720x576)	1620	3240

Memory Complexity

An additional storage for look up table consisting of macro-block number is required for the proposed algorithm. For scalability to CIF (352x288) and higher resolution, the macro-block number needs to be 16bit. The size of the array should be equal to the size of that resolution in macro-blocks. So if the resolution is *(w x h)* then the additional storage requirement is = *w x h * 2 / (16x16)* bytes. The theoretical calculation for additional memory requirement for the proposed algorithm for different video resolutions is given in Table 4.10. As seen from the table, the % increase in the memory overhead is negligible.

COMPRESSION VS. VIDEO QUALITY PERFORMANCE

Considering the process of encryption, the proposed algorithm has almost no effect on compression ratio and also the video quality expressed in terms of Peak-Signal-to-Noise Ratio (PSNR). This can be proved by comparing the size of unencrypted files and encrypted ones as well as comparing the PSNR, as shown in Table 4.11. The parameters chosen for H.264 encryption are: *qp* = 28, interval of IDR frame = 30 and standard video test sequences were taken for experimentation. The average of the result obtained by running the video sequence on 200 frames is taken.

Table 4.11: Video Quality Analysis

Video Sequence	Size / frame (in bytes)		PSNR (of Y component)	
	Without encryption	*With encryption*	*Without encryption*	*With encryption*
Claire	155.125	158.16	39.03	39.03
Foreman	668.975	700.3	35.14	35.16
Hall monitor	264.82	268.705	36.97	36.96

The result shows that there is an increase of nearly 1% in bit rate for the proposed algorithm while keeping the same video quality. There is no such figure available for the reported works. Yuan Li et al [15] had shown that there is nearly 1% increase in bit rate for their algorithm; however they did not give any figure for their video quality.

4.3.4 Discussion

It is shown here that it is possible to incorporate a video encryption algorithm inside the H.264 compression signal chain itself with extremely low computational overhead. Security analysis results are presented to show that it would take considerable effort to break into the encryption system using brute-force attack. The main benefit of the proposed algorithm lies in the fact that in spite of encryption being embedded inside the compression, there is negligible increase in computational complexity and memory requirement and negligible decrease in compression efficiency compared to a traditional H.264 compression system for a given video quality. The results obtained after implementation clearly corroborates these features (Table1 4.9, Table 4.10 and Table 4.11).

4.4 Conclusion

In this chapter a set of novel watermarking and encryption algorithms were proposed that can be used for DRM and access control of security-sensitive video-centric applications of HIP like multimedia content distribution in distance education and patient-doctor video chat in remote medical consultation. The main differentiating feature of both the proposed watermarking and encryption algorithms are their low-computational complexity without compromising on the video quality, while still providing adequate security.

For watermarking, in addition to proposing a novel low-overhead watermark embedding algorithm, the other algorithms required for a complete system like frame level integrity check after embedding the watermark, finding space and location inside the video for embedding the watermark, handling .images and text strings separately etc. were designed and implemented. A modified version of H.264 Intra-prediction mode calculation is also proposed to take care of watermarking requirements. An implementation of watermarking evaluation tool and a novel methodology to evaluate watermark against attacks is also presented. The system is implemented on HIP and results on computational complexity presented to support the claim of keeping the complexity low. It is also shown through experiments on standard video test sequences that the video quality does not deteriorate after embedding the watermark. Finally, a detailed set of experimental results are presented using the tool and methodology introduced to prove the security robustness of the proposed algorithm.

For encryption, a novel two-stage algorithm is proposed that keeps the computational overhead low through doing separate header encryption and through novel use of the flexible macro-block reordering (FMO) feature of the H.264/AVC as the encryption operator. The system is implemented on HIP and results on computational complexity presented to support the claim of keeping the complexity low. Security analysis of the algorithm is also presented to prove their robustness against brute-force attacks. It is also shown through experiments that in spite of being computationally efficient and secure, the proposed algorithm does not deteriorate the video quality.

References

[1]. ITU-T Rec. H.264, "Advanced Video Coding for Generic Audiovisual Services", 2010.

[2]. T. Wiegang, G.J. Sullivan, G. Bjøntegaard, and A. Luthra. "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circuits Syst. Video Technol.*, 13(7), Jul. 2003.

[3]. G. Do¨err, J-L. Dugelay, "A guide tour of video watermarking", *Elsevier Journal of Signal Processing: Image Commun.*, vol. 18 (4), April 2003.

[4]. F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video", *Signal Processing Journal*, vol. 66 (3), Nov. 1998.

[5]. D. Simitopoulos, S.A. Tsaftaris, N.V. Boulgouris, M.G. Strintzis,"Compressed-domain video watermarking of MPEG streams", *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, vol. 1, Aug. 2002.

[6]. Gang Qiu, Pina Marziliano, Anthony T.S. Ho, Dajun He, Qibin Sun, "A hybrid watermarking scheme for H.264/AVC video", *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, August 2004.

[7]. Maneli Noorkami, Russell M. Mersereau, "Compressed-Domain Video Watermarking for H.264", *IEEE International Conference on Image Processing (ICIP)*, Sept. 2005.

[8]. Guo-Zua Wu, Yi-Jung Wang, Wen-Hsing Hsu, "Robust watermark embedding/detection algorithm for H.264 video", *Journal of Electronic Imaging 14(1)*, March 2005

[9]. ITU-T Tutorial, "Objective perceptual assessment of video quality: Full reference television", 2004

[10]. L. Qiao and K. Nahrstedt, "Comparison of MPEG encryption algorithms", *International Journal on Computers & Graphics, Special Issue: "Data Security in Image Communication and Network"*, Vol. 22, No. 3, 1998.

[11]. Tosun, A.S. Feng, W.-C, "Efficient multi-layer coding and encryption of MPEG video streams", *IEEE International Conference on Multimedia and Expo, ((ICME)*, Volume: 1 July 2000.

[12]. Z. Liu and X. L. Sch, "Motion vector encryption in multimedia streaming", *Proc. of 10th International Multimedia Modelling Conference*, Australia, 2004.

[13]. Yao Ye; Xu Zhengquan; Li Wei, "A Compressed Video Encryption Approach Based on Spatial Shuffling", *The 8th International Conference on Signal Processing (ICSP)*, vol.4, Nov. 2006.

[14]. Changgui Shi; Bhargava, B., "An efficient MPEG video encryption algorithm", *Proc. of the Seventeenth IEEE Symposium on Reliable Distributed Systems*, vol., no., Oct 1998.

[15]. Y. Li, L. Liang, Z. Su and Jianguo Jiang, "A New Video Encryption Algorithm for H.264", *Proc. of Fifth International Conference on Information, Communications and Signal Processing (ICICS'05)*, Page(s) 1121-1124, Thailand 2005.

[16]. Y. Zou, T. Huang, W.Gao and L. Huo, "H.264 video encryption scheme adaptive to DRM", *IEEE Transactions on Consumer Electronics*, Vol.52, no.4, Nov. 2006.

5

Context-aware Intelligent TV-Internet Mash-ups

5.1 Introduction

In **chapter 2**, HIP was introduced as a basic device that can make the television connected to the internet world and a user study was presented which indicated the need for seamlessly blending internet experience into TV experience and enrich the standard broadcast TV watching experience through internet capability. This necessity of ubiquitous experience translates into the need for applications that can understand what the user is watching on broadcast TV (termed to as TV-context) and provide user with additional information / interactivity on the same context using the internet connectivity. This kind of applications is termed as TV-Internet mash-up.

Understanding the basic TV context (what channel is being watched and what the is content of the program) is quite simple for digital TV broadcast like IPTV using metadata provided in the digital TV stream [1]. But, in developing countries, IPTV penetration is almost zero and even penetration for other kinds of digital TV like digital cable or satellite direct-to-home (DTH) is also quite low (less than 10% in India). Even for the small percentage of digital cable or satellite DTH coverage, the content is not really programmed to have context-metadata suitable for interactivity as there is no return path for the interactivity. This is mainly due to need for keeping content compatibility to the analog legacy system; additionally cost and infrastructure issues also play a role. HIP, by its inherent internet enabled architecture, has no issues with return path and has capability to blend video with graphics. Hence it is worthwhile to explore possibility of providing context-aware TV-Internet mash-up based applications on HIP. Fig. 5.1 explains the set up in which HIP can be used to provide such applications.

Quite a few interesting applications can be created using context-aware TV-internet mash-ups. A few such applications with novel ways to detect channel identity and textual context are proposed in subsequent sections of 5.2, 5.3 and 5.4. In section 5.2 a novel channel identity detection methodology is introduced. In section 5.3 a novel textual context detection methodology in static pages of video is proposed and in section 5.4 a novel textual context detection methodology in broadcast news video is presented. In each of sections 5.2, 5.3 and 5.4, the problem definition aided by literature survey is discussed, followed by proposed algorithms, systems descriptions, and experimental results with discussion. Finally section 5.5 provides the summary and conclusion.

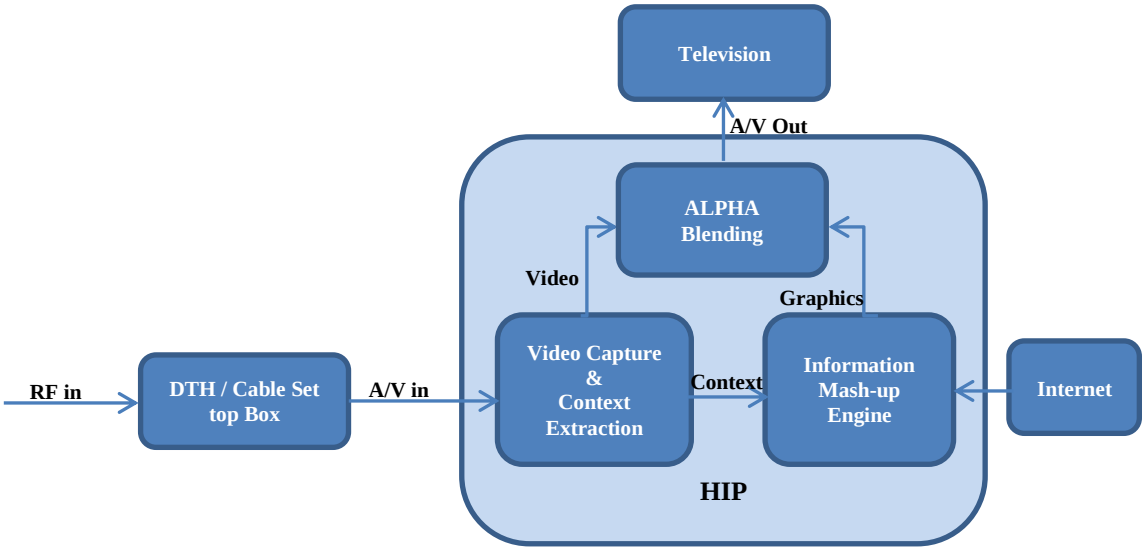


Fig. 5.1: Using HIP for TV-Internet Mash-ups

5.2 TV Channel Identity as Context

5.2.1 Problem Definition

TV-Internet mash-up applications like Electronic Program Guide (EPG), TV audience viewership rating, targeted advertisement through user viewership profiling, social networking among people watching the same program etc. can benefit from identifying which channel is being watched [2]. TV audience viewership rating applications often use audio watermarking and audio signature based techniques to identify the channels [2], [3], [4]. However, audio watermarking based techniques, though real-time, needs modification of the content on the broadcaster end. Audio signature based techniques do not need content modification on the broadcaster end, however they require sending captured audio feature data of channel being watched to back-end for offline analytics and hence cannot be performed in real-time. Since the focus is on broadcaster-agnostic real-time TV-internet mash-up kind of applications, these techniques will not work well. Hence it is needed to look for alternate techniques that should work in real-time and should be computationally lightweight so that it can be run on HIP.

In our proposed work, the possibility of using TV channel logo for channel identification is explored. Each TV channel broadcast shows its unique logo image at pre-defined locations of the screen. The identification can be typically done by doing image template based matching of the unique channel logo. Fig. 5.2 gives a screenshot of the channel logo in a broadcast video of a couple of channels.

Fig. 5.2: Channel Logos in Broadcast Video

Looking at the logos of most popular 110 Indian TV channels it is found that the logos can be can be classified into 7 different types –

a) Static, opaque and rectangular
b) Static, opaque and non-rectangular
c) Static, transparent background and opaque foreground
d) Static, alpha-blended with video
e) Non-static, changing colors with time
f) Non-static, fixed animated logos
g) Non-static, randomly animated logos.

The proportion of the channels classified in these seven types is given in Fig. 5.3. In the requirement of the current work, all the types of channels except the non-static randomly animated ones are considered. The reason for exclusion is the fact that such channels do not have a unique signature to detect and anyway their proportion is also very small (1%).

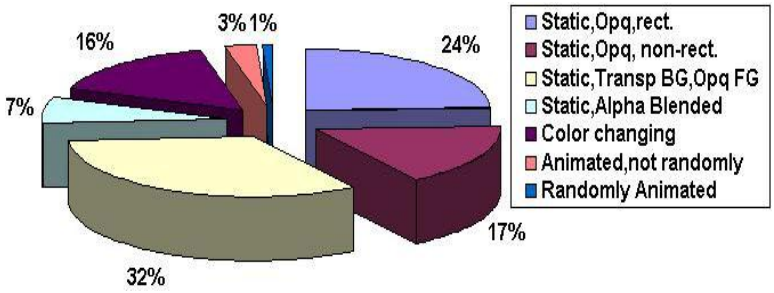


Fig. 5.3: Logo Types

Some related work in this field is described in [5], [6], [7], [8] and [9]. All the approaches were analyzed and it was found that the best performance is observed for the approaches described in [5] and [9]. But the approaches taken in [9] involve Principal Component Analysis (PCA) and Independent Component Analysis (ICA), both of which is very much computationally expensive and thus is difficult to be realized on HIP to get a real time performance. The approach of [5] works well only for channel type (a) – static, opaque and rectangular logos. Hence there is need for developing a channel logo recognition algorithm that on one side should be lightweight enough to be run in real-time on HIP and on the other side should detect all the six types of channel logos considered (type a to type f). There are solutions available in the market like MythTV[13], which provide channel logo based detection features, but it does not support all types of channel logos and it also does not support the SDTV resolution PAL TV standard prevalent in India. The main contribution of the proposed work has been four fold –

a) A design is proposed that reduces the processing overhead by limiting the search space to known positions of logo and integrates an available lightweight color template based matching algorithm to detect logos.
b) A novel algorithm is devised to automatically declare any portion of the logo to be "don't care" in order to take care of the non-rectangular, transparent and alpha-blended static logos (types b, c and d). This makes use of the fact that static portions of the logo will be time-invariant whereas transparent or alpha-blended portions of the logo will be time-varying. It also innovatively applies radar detection theory as a post-processing block to improve the accuracy of the detection under noisy video conditions that are prevalent in analog video scenarios.
c) To make the logo detection work reliably for non-static logos (types e and f), it is proposed to create a sequence of logo templates covering the whole time variation cycle of the logo and to correlate the captured video with the set of templates to find the best match.

[13] www.mythTV.com

d) To save on the scarce computing resources, the logo detection algorithm is not run all the time. The system uses an innovative blue-screen / blank-screen detection during channel change as an event to trigger the logo detection algorithm only after a channel change.

A paper has already been published on this work on logo detection (Appendix B – [5]) and a patent has been filed (Appendix B – [e]).

5.2.2 Proposed System

The overview of channel logo recognition methodology is presented in Fig. 5.4 in the context of the overall system presented in Fig. 5.1. Each step is then subsequently elaborated in detail.

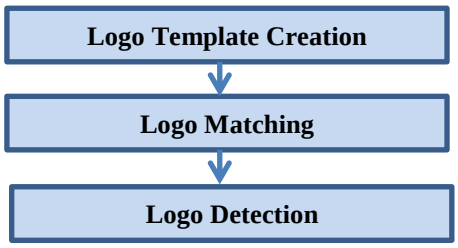


Fig. 5.4: Overview of Channel logo recognition

LOGO TEMPLATE CREATION

Initially the videos of all channels are recorded to create a single video file. Manual annotation is performed on the video file to generate a ground-truth file containing channel code, start frame number and end frame number. This video is played using a tool that enables the user to select the Region of Interest (ROI) of the logo from the video using mouse. To aid the user, a ROI suggestion system is provided in the tool, which is introduced below as an innovative extension as given in (5.1). The tool takes the annotated ground-truth file as input to generate the logo template file containing ROI coordinates, height and width of ROI and a feature-based template for each channel logo. The feature considered for the template generation is quantized HSV values of the pixels in the ROI [5]. To reduce the template size without affecting the detection performance, 36 levels of quantization are taken. It should be noted that input video comes in UYVY format (as per HIP implementation), so the tool coverts this video to HSV.

Algorithm for ROI suggestion

The algorithm is based on the principle that logo region remains invariant midst varying video. The video buffer f_i: contain quantized HSV values of all pixels in i^{th} frame.

- Compute the run-time average of each pixels of i^{th} frame at (x, y) coordinate $a_i(x,y)$ as

$$a_i(x,y) = \frac{(a_{i-1}(x,y) * (i-1) + f_i(x,y))}{i}$$

- Compute dispersion $d_i(x, y)$ of each pixel of i^{th} frame as

$$d_i(x,y) = d_{i-1}(x,y) + abs(a_i(x,y) - f_i(x,y))$$

- Compute variation $v_i(x, y)$ in pixel value at location (x, y) at i^{th} frame as

$$v_i(x,y) = \frac{d_i}{i}$$

- Suggest the pixels having a variance greater than threshold as out of logo region

$$f_i(x,y) = DON'TCARE \forall x,y \in v_i(x,y) > Th_{var} \quad \textbf{(5.1)}$$

LOGO MATCHING

The template of each logo is uploaded to the HIP box. Inside the box, the captured TV video in the corresponding ROI is compared with the template using Correlation Coefficient based approach. The score always gives a value in the range of 0 to 1. The logo is considered as a candidate if the score is greater than a fixed threshold. For noise-free videos, a fixed threshold arrived at using experimentation and heuristics works well. However, for noisy videos, one needs to go for statistical processing based decision logic. Usually first the fixed threshold based algorithm is applied with threshold kept on the lower side (0.75 in our case) to arrive at a set of candidate channels with best matching scores. This normally contains quite a few false positives. The standard M/N detection approach used in Radar Detection theory [10] is employed to reduce the false positives. The logo scores are generated for every f frames of video, where f is the averaging window length. A decision algorithm is implemented using N consecutive scores. The channel that is occurring at least M times out of N is detected as the recognized channel. The optimal value of M and N are found experimentally to be 5 and 9 respectively.

For time-varying logos at fixed locations (logo types e and f), it is observed that the variation follows a fixed pattern over time. It is seen that either the color of the logo goes through a cycle of variation or the image of the logo itself is animated going through a fixed animation cycle. For both these cases, instead of taking one image of the logo as template a series of images of the logo (representing its full variation cycle either in color level or in image level) is taken as a template set and same methodology as proposed above followed by some aggregation logic is used.

Logo detection is a resource hungry algorithm as it does pixel by pixel matching for correlation. Hence it should be triggered only when there is a channel change. The change in channel is detected using the blue or back screen that comes during channel transitions. In the proposed system, it runs logo detection every 15 seconds until a channel is detected. Once detected the next logo detection is triggered only by a channel change event. This frees up useful computing resource on HIP during normal channel viewing which can be used for textual context detection proposed in sections 5.3 and 5.4.

5.2.3 Results

The channel logo recognition module is tested with videos recorded from around 110 Indian channels. The accuracy of recognition is measured using two parameters namely recall and precision Recall *(r)* and precision *(p).*

$$r = \frac{c}{c+m}, p = \frac{c}{c+fp} \quad \textbf{(5.2)}$$

Where, c is number of correct detections, m is number of misses and fp is the total number of false positives.

For the 110 channels tested, it is found to have $r = 0.96$ and $p = 0.95$. As is seen from the results, the accuracy of the algorithm is quite good. The reasons for the small recall and precision inaccuracy can be explained as follows -

- The channel logos with very small number of pixels representing the foreground pixel of the channel logo are missed in 1% cases.
- The reason behind the misses for rest 3% cases is that the channel logo is shifted to a different location from its usual position or channel logo itself has changed. A sample screen shot of the channel logo location shift in Ten Sports channel is shown in Fig. 5.5. Sample screenshots of the channel logo color change in Sony Max channel and altogether change in channel logo for Star Plus channel is showed in Fig. 5.6 and Fig. 5.7 respectively.

To explain the 5% false positive results, the details are presented in form of a confusion matrix in the Table 5.1. It is evident that most of the channels are mainly confused with DD Ne. The major

reason behind it is that DD Ne channel logo is very small in size and false positives can be improved by removing DD Ne template from the corpus. The reason for Zee Punjabi and Nepal-1 being detected wrongly is because these logos are transparent and false detection occurs in some conditions of the background video. It does not happen all the time and hence can be improved through time averaging.

Fig. 5.5: Channel Logo Location Shift

Fig. 5.6: Channel Logo Color Change

Fig. 5.7: Channel Logo Image Change

Table 5.1 : Confusion Matrix for Channel Logo Recognition

Original Channel	Detected As
Zee Trendz	DD Ne
Zee Punjabi	TV9 Gujarati
DD News	DD Ne
Nick	DD Ne
Nepal 1	Zee Cinema

The computational complexity of the proposed system was also measured and the results are shown in Table 5.2 for different parts of the algorithm. As is seen from the results, the system is able to detect the channel at less than 1.5 seconds after the channel change which is quite acceptable from the user experience perspective. However since logo detection is triggered by channel change, the DSP CPU is available for other tasks when the user is not changing the channels.

Table 5.2 Time complexity of Different Algorithm Components

Module	Time (msec)
YUV to HSV	321.09
ROI mapping	0.08
Mean SAD matching	293.65
Correlation	847.55

5.2.4 Discussion

A logo recognition based channel identification technique was proposed here for value-added TV-Internet mash up applications. For logo recognition, a solution using template-based matching is introduced, where the logo templates are generated offline and the logo recognition is performed on the captured TV video in real-time on HIP boxes using the templates. The main contribution of the proposed work has been four fold –

a) An algorithm to suggest Logo ROI during manual template generation

b) Algorithm to handle the non-rectangular, transparent and alpha-blended static logos with improved detection accuracy using statistical decision logic
c) Time sequence based algorithm to handler non-static logos
d) Channel change event detection as trigger to logo recognition algorithm for reduced computational overhead

Results of experimental study of 110 Indian TV channels are presented. Results show a recall rate of 96% and precision rate of 95%, which is quite acceptable. The cases where the algorithm is failing is analyzed and is found that the failures can be explained specific conditions that can be handled in specialized manner and is kept as a scope for future work. The time complexity of the algorithm is also profiled and it is found that a channel can be detected within 1.5 seconds of a channel change.

5.3 Textual Context from Static Pages in Broadcast TV

5.3.1 Problem Definition

Active services are value added interactive services provided by the DTH (Direct-To-Home) providers and are designed based on digital video broadcasting standard for satellites (DVB-S). They include education, sports, online banking, shopping, jobs, matrimony etc. These services provide interactivity by using Short Messaging Service (SMS) as the return path. For instance, the consumers can interact by sending an SMS having a text string displayed on the TV screen to a predetermined mobile number. For example, Tata Sky, the leading DTH provider in India[14] provides services like Active Mall to download wall papers and ringtones, Active Astrology, Active Matrimony, Movies on demand, Service Subscription, Account Balance etc.

As the return path for the traditional DTH boxes are not available, as part of interactivity, these pages instruct the user to send some alphanumeric code generated on the TV screen via SMS from their registered mobiles, This is illustrated in Fig. 5.8 with a screenshot of an active page of Tata Sky, with the text to SMS marked in red. The system is quite cumbersome form user experience perspective. A better experience can be provided if the texts in the video frame can be recognized automatically and SMS is generated. There was no work found in the literature on text recognition in static TV video frames.

In our proposed system, an optical character recognition based approach is presented to extract the SMS address and text content from video to send SMS automatically by just pressing a hot key in the HIP remote control. In addition to the complete end-to-end system implementation, the main contribution is in the design of an efficient pre-processing scheme consisting of noise removal, resolution enhancement and touching character segmentation, after which standard binarization techniques and open source print OCR tools like GOCR[15] and Tesseract[16] are used to recover and understand the textual content. There are OCR products like Abbyy Screenshot Reader and Abbyy FineReader[17] also available; however, it was decided to use open source tools to keep system cost low.

A patent has been filed on this work on text OCR in static TV pages (Appendix B- [f]).

5.3.2 Proposed System

The proposed system is implemented using the generic architecture given in Fig. 5.1. After collecting a set of images of active pages, the following observations can be made:

[14] www.tatasky.com
[15] http://jocr.sourceforge.net/
[16] http://sourceforge.net/projects/tesseract-ocr/
[17] http://www.abbyy.com/

- The location of the relevant text region is fixed for a particular active page
- There is a considerable contrast difference between the relevant text region and the background
- The characters to be recognized are of standard font type and size

Fig. 5.8: Text Embedded in Active Pages of DTH TV

Based on these observations, a set of steps for the text recognition algorithm as depicted in Fig. 5.9 is proposed. Each of the steps is subsequently elaborated in detail.

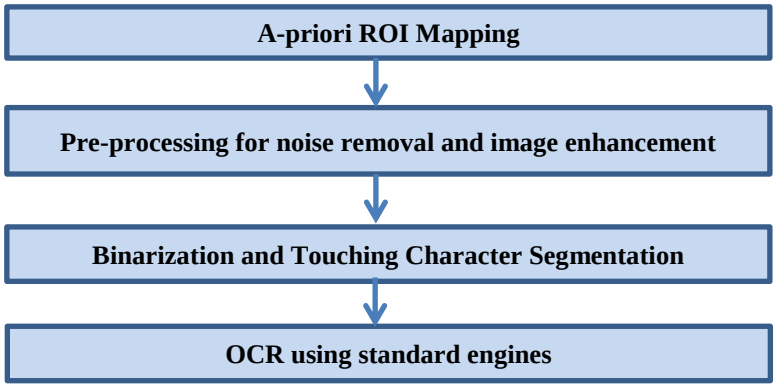


Fig. 5.9: Text Recognition is Static Pages

A-PRIORI ROI MAPPING

In this phase the relative position of all relevant text region for each active pages are manually marked and stored in a database. First the bounding box coordinates for each ROI is found in the reference active pages through manual annotation. This manually found ROI can be used as a-priori information as it was found that the active pages are static.

PRE-PROCESSING

Once the ROI is defined manually one can directly give this ROI to the recognition module of some OCR engine. However it is found that there are a lot of blurring and artifacts in the ROI that reduces the recognition rate of the OCR. Hence a pre-processing scheme is proposed to improve the quality of the text image before giving it to a standard OCR engine for recognition. The pre-processing scheme is divided into two parts – noise removal and image enhancement. For noise removal, a 5 pixel moving window average for the Luminance (Y) values is used. The image is enhanced using the following steps -

- Apply six tap interpolation filter with filter coefficients (1, -5, 20, 20, -5,1) to zoom the ROI two times in height and width

- Apply frequency domain low-pass filtering using DCT on the higher resolution image

ICA based approach can also produce very good result but the above approach is chosen to keep the computational complexity low.

BINARIZATION AND TOUCHING CHARACTER SEGMENTATION

The output of the preprocessed model is then binarized using an adaptive thresholding algorithm. There are several ways to achieve binarization so that the foreground and the background can be separated. However, as both the characters present in the relevant text region as well as the background are not of a fixed gray level value, adaptive thresholding is used in this approach for binarization. To obtain the threshold image, the popular Otsu's method [11] is used.

Once the binarized image is obtained, it is observed quite frequently that the image consists of a number of touching characters. These touching characters degrade the accuracy rate of the OCR. Hence the Touching Character segmentation is required to improve the performance of the OCR. An outlier detection based approach is proposed here, the steps of which are as below -

- Find the width of each character. It is assumed that each connected component with a significant width is a character. Let the character width for the i^{th} component be WC_i
- Find average character width $\mu_{WC} = 1/n \sum_{i=1}^{n} WC_i$ where n is the number of character in the ROI
- Find the Standard Deviation of Character Width (σ_{WC}) as $\sigma_{WC} = STDEV(WC_i)$
- Define the threshold of Character Length (T_{WC}) as $T_{WC} = \mu_{WC} + 3\sigma_{WC}$
- If $WC_i > T_{WC}$ mark the i^{th} connected component as candidate character

AUTOMATIC DETECTION OF THE TEXT BY THE OCR ENGINE

The properly segmented characters obtained as output of the previous module is passed to two standard OCR engines – GOCR and Tesseract for automatic text detection. Once the text is detected, it is automatically sent as SMS to the satellite DTH service provider.

5.3.3 Results

The different kinds of videos are recorded from different kind of DTH active pages available. The screenshots of 10 different frames (only the relevant text region or ROI) is given in Fig. 5.10 (a) to (j). The page contents are manually annotated by storing the actual text (as read by a person) along with the page in a file. The captured video frames are passed through the proposed algorithm and its output (text strings) are also stored in another file. The two files are compared for results.

The performance is analyzed by comparing the accuracy of the available OCR engines (GOCR and Tesseract) before and after applying the proposed image enhancement techniques (Pre-processing, Binarization and Touching Character Segmentation). The raw textual results are given in Table 5.3. The accuracy is calculated from the raw text outputs using character comparison and is presented graphically in Fig. 5.11. From the results it is evident that considerable improvement (10% in average, 50% in some cases) is obtained in character recognition after using the proposed methodology of restricting the ROI and applying preprocessing and touching character segmentation before providing the final image to the OCR engine. It is also seen that Tesseract performs better as an OCR engine compared to GOCR.

5.3.4 Discussion

Here an end-to-end system solution to automate user interaction in DTH active pages is proposed. It works by extracting the textual context of the active page screens through text recognition. In addition to the complete end-to-end system implementation, the main contribution is in the design of an efficient pre-processing scheme consisting of noise removal, resolution enhancement and touching character segmentation, on which standard binarization techniques (like Otsu's) and open source print

OCR tools like GOCR and Tesseract are applied. From the results it is quite clear that the proposed pre-processing schemes improve the text recognition accuracy quite significantly. Additionally it is seen that Tesseract OCR performs much better than GOCR.

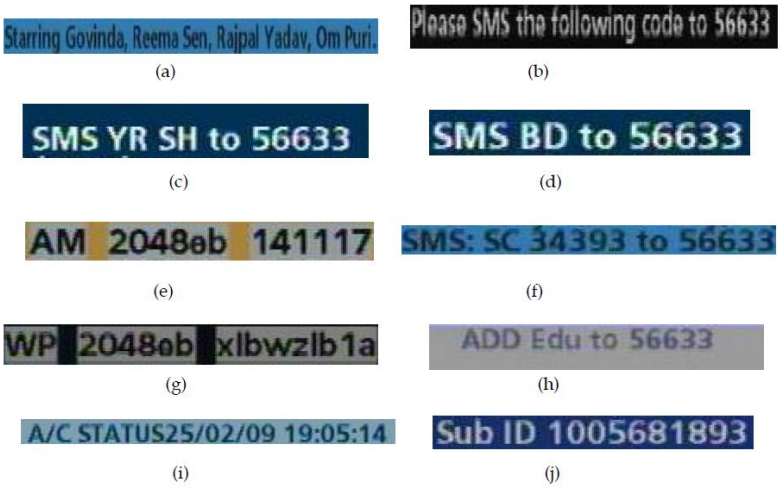


Fig. 5.10: Different Active Page Screenshots (a) – (j)

Table 5.3: Raw Text Outputs from algorithms for different Active Pages

Image	Output of GOCR	Output of Tesseract	After Applying Proposed Algorithms	
			GOCR	Tesseract
(a)	Sta_ring Govind_. Reem_ _n. RajpaI Yadav. Om Puri.	Starring Guvinda, Rcema Sen, Raipal Yadav, Om Puri.	Starring Govind_. Reem_ _n. RajpaI Yadav. Om Puri.	Starring Guvinda. Reema Sen, Raipal Yadav. Om Puri.
(b)	_____ ___ ___ _____ ____ __ __	Pluww SMS thu fnlluwmg (adn In 56633	___ SMS th_ folIcmng cod_ to S__	Planta SMS tha Iullmmng mda tn 56633
(c)	SmS YR SH to	SMS YR SH in 56633	SmS YR SH to _____	SMS YR SH to 56533
(d)	_m_ BD to _____	SMS BD to 56633	SMS BD to S____	SMS BD to 56633
(e)	AM t___o_,_b _q____	AM 2048eb 141117	AM tOa_gb _q____	AM 2048eb 141117
(f)	_M_= _ _A___ to Sd___	SMS: SC 34393 tn 56533	_M_= _ _A___ to Sd___	SMS: SC34393 tn 56633
g)	_W _ ' _b _ Ib_lb _a	W6.} 048abl;lbwzIb1a	___ __Y_b yIbw_Ib_a	WP 2048ab Mlbwzlb 1 a
(h)	ADD Ed_J to S____	ADD Eau to $6633	ADD Ed_J to S____	ADD Edu to 56633
(i)	AIC STAlUSlS/OUO_ t_;OS;t_	AIC STATUS25/02/09 1 9:05:1 4	mlC S_ATUSlS/OUO_ t_;OS=tA	A/C STATUS 25/02/09 1 9:05:14
(j)	_ _____'__	Sub ID 1005681893	WbID_OOS_B_B__	Sub ID 1005681893

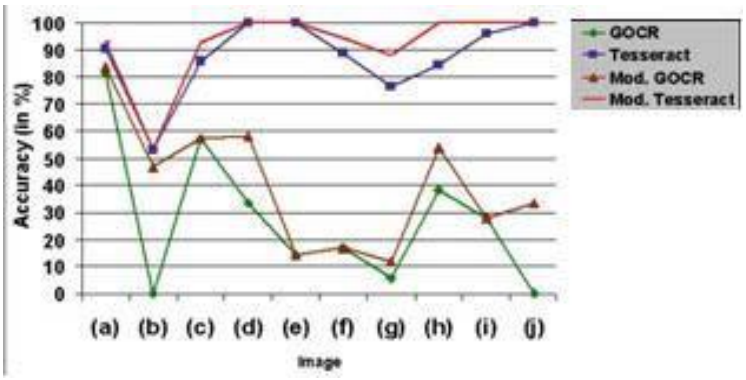


Fig. 5.11: Performance of OCR Engines before and after the Proposed Algorithms

5.4 Textual Context from Text Embedded in Broadcast TV

5.4.1 Problem Definition

Typically broadcast videos of news channels, business channels, music channels, education channels and sport channels carry quite a bit of informative text that are inserted / overlaid on top of the original videos. If this information can be extracted using optical character recognition (OCR), related information from web can be mashed-up either with the existing video on TV or can be pushed into the second-screen devices like mobile phone and tablets. Fig. 5.12 gives an example screenshot of textual information inserted in a typical Indian news channel.

Fig. 5.12: Contextual Text Embedded in TV Video

Gartner report suggests that there is quite a bit of potential for new Connected TV widget-based services[18]. The survey on the wish list of the customers of connected TV shows that there is a demand of a service where the user can get some additional information from Internet or different RSS feeds, related to the news show the customer is watching over TV. A comprehensive analysis on the pros and cons of the products on Connected TV can be found in [12]. But none of the above meets the contextual news mash-up requirement. A nearly similar feature is demonstrated by Microsoft in International Consumer Electronics Show (CES) 2008 where the viewers can access the contents on election coverage of CNN.com while watching CNN's television broadcast, and possibly even participate in interactive straw votes for candidates[19]. But this solution is IPTV metadata based and hence does no need textual context extraction.

The main technical challenge for creating the solution lies in identifying the text area that changes dynamically against a background of dynamically changing video. The state of the art shows that the approaches for text localization can be classified broadly in two types - (i) Using pixel domain information when the input video is in raw format and (ii) using the compressed domain information when the input video is in compressed format. Since the raw (UYVY) video is already being captured as the input for the proposed system, only the pixel domain methods are considered. A comprehensive survey on text localization is described in [13] where all different techniques in the literature from 1994 to 2004 have been discussed. It is seen that the pixel domain approach are mainly Region based (RB) and RB based approached are further sub-divided into Connected Component based (CB) and Edge based (EB). CB based approaches are covered in [14], [15], [16]. [17], [18]. EB based approached are covered in [19], [20], [21], [22], [23]. In [24], [25], [26] one gets combined CB and EB based approaches, whereas [27], [28] combines compressed domain and pixel domain information along with combination of CB and EB methods. It is typically seen that it is difficult to have one

[18] http://blogs.gartner.com/allen_weiner/2009/01/09/ces-day-2-yahoosconnected-tv-looks-strong
[19] http://www.microsoft.com/presspass/press/2008/jan08/01-06MSMediaroomTVLifePR.mspx

particular method perform well against varying kind of texts and video backgrounds - hybrid approaches proposed in [27] and [28] seem to perform well in these scenarios.

In this work, an end to end system is proposed that can provide these features on HIP. The main contribution of the work lies in proposing low-computational-complexity algorithms for

a) An improved method for localizing the text regions of the video and then identifying the screen layout for those text regions, extending the work in [27] and [28].
b) Recognizing the content for each of the regions containing the text information using novel pre-processing techniques and Tesseract OCR as was done in section 5.3.
c) Applying heuristics based key word spotting algorithm where the heuristics are purely based on the observation on the breaking news telecasted in Indian news channels.

Three papers have already been published on this work on text OCR in dynamic TV pages (Appendix B – [6], [7], [13]) and a patent has been filed (Appendix B – [g]).

5.4.2 Proposed System

The proposed system follows the system design presented in Fig. 5.1 and consists of steps given in Fig. 5.13. Each of the steps is presented in detail subsequently.

LOCALIZATION OF THE SUSPECTED TEXT REGIONS

The approach of text localization proposed in [26] is used here also. Our proposed methodology is based on the following assumptions based on the observation from different news videos -

- Text regions have a high contrast
- Texts are aligned horizontally
- Texts have a strong vertical edge with background
- Texts of Breaking news persists in the video for at least 2 seconds

Following [26], first the low-contrast components are filtered out based on intensity based thresholding (output $V_{cont.}$). Then for final text localization, a low-computational complexity algorithm that can localize the candidate regions efficiently is proposed. The methodology is presented as below:

- Count the number of Black pixels in a row in each row of V_{cont}. Let the number of Black pixels in i^{th} row be defined as $cnt_{black}(i)$
- Compute the average (avg_{black}) number of Black pixels in a row as

$$\boldsymbol{avg_{black} = \sum_{i=1}^{ht} cnt_{black}(i)/ht}$$

  where ht is the height of the frame.
- Compute the absolute variation $av(i)$ in number of black pixels in a row from avg_{black} for each row as

$$\boldsymbol{av(i) = abs(cnt_{black}(i) - avg_{black})}$$

- Compute the average absolute variation (aav) as

$$\boldsymbol{aav = \sum_{i=1}^{ht} av(i)/ht}$$

- Compute the threshold for marking the textual region as

$$\boldsymbol{TH_{txt\_reg} = avg_{black} + aav}$$

- Mark all pixels in i^{th} row of V_{cont} as white if

$$cnt_{black}(i) < TH_{txt\_reg} \quad (5.3)$$

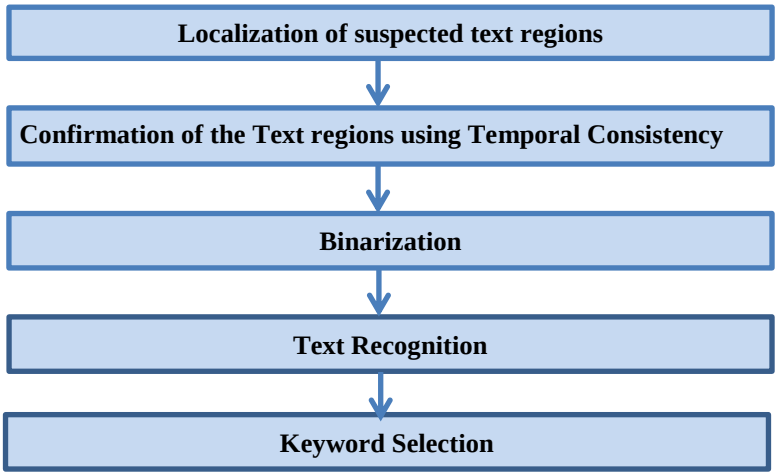


Fig. 5.13: Text Recognition is Broadcast Video

CONFIRMATION OF THE TEXT REGIONS USING TEMPORAL CONSISTENCY

This portion of the proposed method is based on assumption that texts in the breaking news persist for some time. V_{cont} sometime contains noise because of some high contrast regions in the video frame. But this noise usually comes for some isolated frames only and is not present in all the frames in which the breaking news text is persistent. In a typical video sequence with 30 FPS, one frame gets displayed for 33 msec. Assuming breaking news to be persistent for at least 2 seconds, all regions which are not persistently present for more than 2 seconds can be filtered out.

BINARIZATION

Once the pre-processing is done, the vertical and horizontal energy of the sub block is computed based on the assumption that the blocks with text have high energy levels. The regions with lower energy are marked as black after they are checked using a threshold value. First the histogram is computed for all the energy levels in a row, followed by determination of the two major peaks denoting start and end of a text segment. Finally the threshold is taken slightly lower than the smaller peak. The result obtained contains some false positives i.e. noise along with the text detected. Hence morphological operations and filtering are used to enhance the image and give better localization with less false positives. The final rectangular binarized image of the localized text region is fed into the Text Recognition block.

TEXT RECOGNITION

For text recognition, the process outlined in section 5.3 under “Touching Character Segmentation” and “Optical Character Recognition” is followed. The Tesseract OCR engine is used.

KEYWORD SELECTION

Here an innovative post-processing approach is proposed on the detected text based on following observed properties -

- Breaking news are always comes in Capital Letter.
- Font size of breaking news is larger than that of the ticker text
- They tend to appear on the central to central-bottom part of the screen.

These assumptions can be justified by the screen shots of News shows telecasted in different news channels as shown in Fig. 5.14.

From these observations the following approach can be used to identify the keywords -

- Operate the OCR only in upper case

- If the number of words in a text line is above a heuristically obtained threshold value they are considered as candidate text region.
- If multiple such text lines are obtained, chose a line near the bottom
- Remove the stop words (like a, an, the, for, of etc.) and correct the words using a dictionary.
- Concatenate the remaining words to generate the search string for internet search engine - selected keyword can be given to Internet search engines using Web APIs to fetch related news, which can be blended on top of TV video to create a mash up between TV and Web. Since search engines like Google already provide word correction, thereby eliminating the requirement of dictionary based correction of keywords.

5.4.3 Results

The system was tested against 20 news channels, with a number of video sequences from each channel, each of approx. 5 minutes duration. The experimental results are presented and analyzed as below.

ACCURACY OF TEXT LOCALIZATION

A typical video frame and the high contrast region extracted from the frame are shown in Fig. 5.15. In Fig. 5.16, the improved screenshots for the text localization after noise cleaning using the proposed methodology is shown. Referring to the recall and precision measures outlined in (5.3), experimental results show a recall rate of 100% and precision of 78% for the text localization module. The reason behind a low precision rate is tuning the parameters and threshold values in a manner so that the probability of false negative (misses) is minimized. The final precision performance can be only seen after applying text recognition and keyword selection algorithms.

ACCURACY OF TEXT RECOGNITION

Once the text regions are localized each candidate text rows undergo some processing prior to OCR and are given as input to Tessaract for OCR. It is found that in case of false positives a number of special characters are coming as output of OCR. So the candidate texts having special character/ alphabet ratio > 1 are discarded. Moreover proposed keyword detection method suggests that concentrating more on capital letters. So only the words in all capitals are kept under consideration. It is found that character level accuracy of the selected OCR for those cases in improves to 86.57%.

ACCURACY OF INFORMATION RETRIEVAL

Limitations of the OCR module can be overcome by having a strong dictionary or language model. But in the proposed method this constraint is bypassed as the Google search engine itself has one such strong module. So one simply gives the output of OCR to Google search engine and in turn Google gives the option with actual text as shown in Fig. 5.17. The input given to Google was "MUMBAI ATTAC**H**ED" as it is the text detected by the OCR and Google itself gave the corrected text "MUMBAI ATTAC**K**ED" as an option in their "Did you mean" tab. This can be done programmatically using web APIs provided by Google.

Finally in Fig. 5.18, a screenshot of the final application is presented, where the "Mumbai Attacked" text phrase identified using the proposed system is used to search for relevant news from Internet and one such news ("The Taj Attack took place at 06:00 hrs") is superposed on top of the TV video using alpha blending in HIP.

5.4.4 Discussion

In this section, an end to end system on HIP is proposed that provides low-computational-complexity algorithms for text localization, text recognition and keyword selection leading towards a set of novel context-aware TV-Web mash-up applications. As seen from the results, the proposed pre-processing algorithms for text region localization in TV news videos gives pretty good accuracy

(~87%) in final text recognition, which when used with word correction feature of Google, gives almost 100% accuracy in retrieving relevant news from the web.

Fig. 5.14: Screen shots showing breaking news in four different channels

Fig. 5.15: High Contrast Regions in the Video

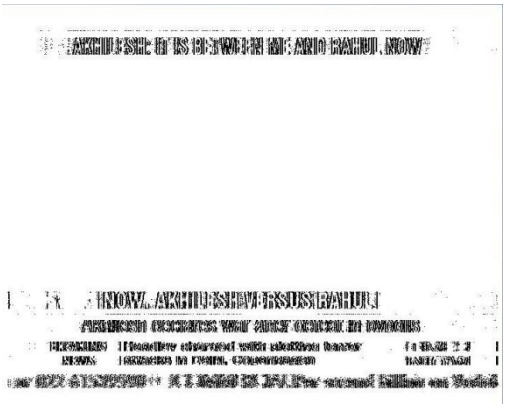


Fig. 5.16: Text Regions after Noise Cleaning

Fig. 5.17: Screen shot of the Google Search Engine with Recognized Text as Input

Fig. 5.18: Screen shot of the Final Application with TV-Web Mash-up

Finally it is shown how this information retrieved from web can be mashed up with the TV video using alpha blending on HIP. As a scope of future work, the same information also can be displayed on the second screen of the user like mobile phones and tablets. There is also scope of working on Natural Language Processing (NLP) for regional news channels and giving cross-lingual mash-ups.

5.5 Conclusion

In this chapter, a novel system of mashing up of related data from internet is presented. It is done by understanding the broadcast video context. Three different novel methodologies for identifying TV video context has been presented–

- Low-computational complexity channel identification using logo recognition and using it for an web-based fetching of Electronic Program Guide for analog TV channels
- Detecting text in static screens of Satellite DTH TV active pages and using it for an automated mode of interactivity for the end user. Text detection accuracy is improved using novel pre-processing techniques
- Detecting text in form of breaking news in news TV channels and using it for mashing up relevant news from the web on TV. Text detection accuracy is improved using novel text localization techniques and computational complexity is reduced using innovative methodologies utilizing unique properties of the "Breaking News" text and using search engine text correction features instead of local dictionary.

Experimental results show that the applications are functional and work with acceptable accuracy. For developing nations, this is the best way to bring power of internet to masses, as the broadcast TV

medium is still primarily analog and the PC penetration is very poor. It is a mean to improve the poor internet browser interactivity reported in the HIP user study (**Chapter 2**).

References

[1]. ITU-T Technical Report, "Access to Internet-sourced contents", *HSTP-IPTV-AISC (2011-03),* March 2011

[2]. M. Fink, M. Covell, S. Baluja (2006). "Social- and Interactive-Television Applications Based on Real-Time Ambient-Audio Identification", *Proceedings of EuroITV.* 2006.

[3]. Shumeet Baluja, Covell, M., "Content Fingerprinting Using Wavelets", *3rd European Conference on Visual Media Production, (CVMP 2006),* Nov. 2006, London.

[4]. Shumeet Baluja, Covell, M.,, "Waveprint: Efficient wavelet-based audio fingerprinting", *Elsevier: Pattern Recognition,* Volume 41, Issue 11, November 2008.

[5]. T. Chattopadhyay, and Chandrasekhar Agnuru, "Generation of Electronic Program Guide for RF fed TV Channels by Recognizing the Channel Logo using Fuzzy Multifactor Analysis", *International Symposium on Consumer Electronics (ISCE 2010),* June 2010, Germany.

[6]. E. Esen, M. Soysal, T. K. Ates, A. Saracoglu, A. Aydin Alatan, "A fast method for animated TV logo detection", *CBMI 2008.* June 2008.

[7]. Ekin, A.; Braspenning, E.,"Spatial detection of TV channel logos as outliers from the content", in *Proc. VCIP. SPIE,* 2006.

[8]. J Wang, L Duan, Z Li, J Liu, H Lu, JS Jin, "A robust method for TV logo tracking in video streams", *ICME,* 2006.

[9]. N.Ozay, B. Sankur, "Automatic TV Logo Detection And Classification In Broadcast Videos", *EUSIPCO 2009,* Scotland, 2009.

[10]. Merill I. Skolnik, "Introduction to Radar Systems", ISBN-10: 0072881380, *McGraw-Hill* 3rd Edition, 2002.

[11]. N. Otsu,, "A threshold selection method from gray-level histograms", *IEEE Trans. Systems, Man, and Cybernetics,* vol. 9, no. 1, 1979.

[12]. Harry McCracken, "The Connected TV: Web Video Comes to the Living Room", *PC World,* Mar 23, 2009.

[13]. K. Jung, K. I. Kim, and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", *Pattern Recognition,* Volume 37, Issue 5, May 2004.

[14]. P. Shivakumara, Q. P. Trung, L. T. Chew; "A Gradient Difference Based Technique for Video Text Detection", *Proceedings of 10th International Conference on Document Analysis and Recognition,* 2009, 26-29 July 2009.

[15]. P. Shivakumara, T.Q. Phan, T. C. Lim, "A Robust Wavelet Transform Based Technique for Video Text Detection", *Proceedings of 10th International Conference on Document Analysis and Recognition (2009),* 26-29 July 2009

[16]. C. Emmanouilidis, C. Batsalas, N. Papamarkos, "Development and Evaluation of Text Localization Techniques Based on Structural Texture Features and Neural Classifiers", *Proceedings of 10th International Conference on Document Analysis and Recognition,* 2009, pp.1270-1274, 26-29 July 2009.

[17]. Y. Jun, H. Lin-Lin, L. H. Xiao, "Neural Network Based Text Detection in Videos Using Local Binary Patterns", Proceedings of Chinese Conference on Pattern Recognition, 2009, pp.1-5, 4-6 Nov. 2009.

[18]. J. Zhong, W. Jian; S. Yu-Ting, "Text detection in video frames using hybrid features", *Proceedings of International Conference on Machine Learning and Cybernetics (2009),* 12-15 July 2009.

[19]. C.-W. Ngo, C.-K. Chan, "Video text detection and segmentation for optical character recognition", *Multimedia Systems,* vol.10, No.3, Mar 2005.

[20]. M. Anthimopoulos, B. Gatos, I. Pratikakis, "A Hybrid System for Text Detection in Video Frames", *Proceedings of The Eighth IAPR International Workshop on Document Analysis Systems (2008),* 16-19 Sept. 2008.

[21]. P. Shivakumara, T.Q. Phan, T. C. Lim, "Video text detection based on filters and edge features", *Proceedings of IEEE International Conference on Multimedia and Expo (2009),* June 28 – July 3 2009.

[22]. P. Shivakumara, T.Q. Phan, T. C. Lim, "Efficient video text detection using edge features", *Proceedings of 19th International Conference on Pattern Recognition (2008),* 8-11 Dec. 2008.

[23]. P. Shivakumara, T.Q. Phan, T. C. Lim, "An Efficient Edge Based Technique for Text Detection in Video Frames", *Proceedings of The Eighth IAPR International Workshop on Document Analysis Systems (2008),* 16-19 Sept. 2008.

[24]. S. Yu, W. Wenhong, "Text Localization and Detection for News Video", *Proceedings of Second International Conference on Information and Computing Science (2009),* 21-22 May 2009.

[25]. Y. Su, Z. Ji, X. Song, R. Hua, "Caption text location with combined features using SVM", *Proceedings of 11th IEEE International Conference on Communication Technology (2008),* 10-12 Nov. 2008.

[26]. Y. Su, Z. Ji, X. Song, R. Hua, "Caption Text Location with Combined Features for News Videos", *Proceedings of International Workshop on Geoscience and Remote Sensing and Education Technology and Training (2008),* 21-22 Dec. 2008.

[27]. T. Chattopadhyay, Aniruddha Sinha, "Recognition of Trademarks from Sports Videos for Channel Hyperlinking in consumer end", Proc. of the 13th International Symposium on Consumer Electronics (ISCE'09), 25-28 May, Japan, 2009.

[28]. T. Chattopadhyay, Ayan Chaki, "Identification of Trademarks Painted on Ground and Billboards using Compressed Domain Features of H.264 from Sports Videos," National Conference on Computer Vision Pattern Recognition, Image Processing and Graphics (NCVPRIPG), January 2010, Jaipur, India.

6
Novel On-screen Keyboard

6.1 Introduction

As deduced from the HIP user study in **Chapter 2**, it has been a challenge to provide cost-effective and easy-to-use text-entry mechanism for accessing services like internet, email and short message services (SMS) from television. This arises mainly from the fact that TV viewing is normally done at distance and hence needs wireless keyboard for text entry. However, a full-fledged separate wireless keyboard (Bluetooth or RF) adds up significantly to the cost. A more affordable option is using infra-red remotes with on-screen keyboard on TV screen.

However, as seen from the user study in **Chapter 2**, traditional "QWERTY" on-screen keyboards require a large number of keystrokes to navigate which makes it cumbersome to use. Hence there is a need for an on-screen keyboard on HIP which allows user to easily navigate and select characters with reduced number of key-strokes, thereby making it more user-friendly. There is also need for doing adequate user studies to prove the efficacy of any such new system proposed from the user experience perspective.

In section 6.2, the problem statement is introduced backed by state-of-the-art study and gap analysis in the context of HIP. In section 6.3, a novel on-screen keyboard layout based on hierarchical navigation is presented along with a mathematical algorithm to arrive at an optimal layout. In section 6.4, user study methodologies based on both heuristic and industry-standard models are presented along with results to show the efficacy of the proposed on-screen keyboard layout. Finally summary and conclusion is provided in section 6.5.

6.2 Problem Definition

The people who use the above mentioned services over the TV may not have previous exposure to computer and keyboard. So the interface or mechanism to use the services should not be complicated, daunting or intimidating. They should be presented with a familiar and encouraging interface to interact with the services on their TV. People use remote controls naturally when using their TV. So the ability to use the remote control for the class of Infotainment services offered in HIP keeps user in natural, familiar and comfortable frame of mind while at the same time, it allows them to enjoy the benefits of these new breed of convergent services. There are user studies presented in [1] and [2],

which also point towards separate user interface requirement for TVs for convergent and interactive services.

Apart from being natural and intuitive, the interaction mechanism should allow user to accomplish their desired task easily and quickly, rather than spending too much time dealing with the interface mechanism. The biggest challenge using remote is text entry, which is normally addressed using on-screen keyboards.. There are proposed on-screen keyboards in literature ([3] to [6]), however none of them address the usability aspect for a non-computer-savvy user. In [3] and [4], on-screen keyboard dynamic layouts are proposed based on frequency of letters and prediction, which more often than not confuses the user. They also do not address the issue of convenient text entry from a remote control. In [5] and [6], the authors present on-screen keyboards that can be used with a mouse / cursor kind of feature on the remote control. Reference [5] also presents some novel cyclic key layout. However, it is observed that due to the limitation of infra-red protocol (RC5 / RC6), on-screen movement of mouse is quite slow resulting in poor user experience. In [7], a user study based comparative analysis is presented, which emphasizes the need for intelligent layouts and text entry mechanisms for remote control based on-screen keyboards. References [8] and [9] caters to specialized applications, and hence not suitable for a generic infotainment device like HIP.

As the main contribution of the work, a novel layout of characters and symbols for the on-screen keyboard is proposed based on a mathematical formulation, which significantly reduces the number of key strokes while typing. After designing the keyboard a set of user studies were conducted to evaluate and monitor the acceptance of this design. Based on the study results some enhancements were incorporated in the layout.

As a more formal methodology for user study evaluation, a design space exploration approach is adopted based on the popular Keystroke-level-model (KLM) and Goal-Operator-Methods (GOMS) model [10] to evaluate the performance of candidate onscreen keyboard layouts. The KLM model gives the low-level description of what users must perform as an operation and the GOMS model gives a structured, multi-level description of the tasks to be performed. The approach enables modeling of user behaviors and actions during a particular task and hence is able to analyze the user experience. Usually KLM performance is pretty good as revealed by applications in [11], [12], [1]3], where the predicted results are very close to actual values obtained through user study. Typically, KLM-GOMS is best suited to evaluate and time specific tasks that require, on average, less than 5 minutes to complete. Our case fits into this requirement.

Subsequent to using KLM-GOMS to formally evaluate the efficacy of the proposed on-screen keyboard layout, as another contribution in this chapter, it is proposed to extend the standard KLM operator set to model the remote based operations and hierarchical layouts. A finger movement operator has been proposed to be included in the KLM, based on [14] where the P operator as used for modeling standard pointing devices, such as mouse, have been experimentally re-estimated, for the case of a remote and a hierarchically blocked layout. This extended KLM can therefore be applied to evaluate other interfaces of similar class as well.

The work done in line with the contributions mentioned above has already resulted in three publications (Appendix B - [8], [9], [15]. One patent has been filed on the on-screen keyboard layout (Appendix B – [h]).

6.3 Proposed System

To achieve the stated goals, a system is proposed where an on-screen keyboard is displayed on the monitor of a television using HIP and is operated by a remote control (Fig. 6.1) which has among other things, has 9 special keys for performing navigation and selection of character or symbols. All the special keys are marked in the figure.

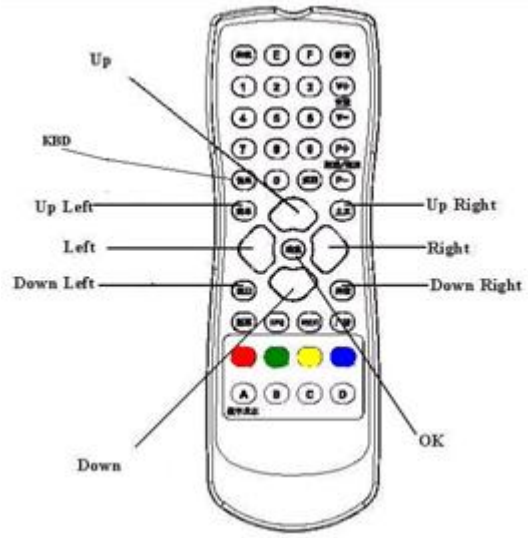


Fig. 6.1: Layout of the Accompanying Remote Control

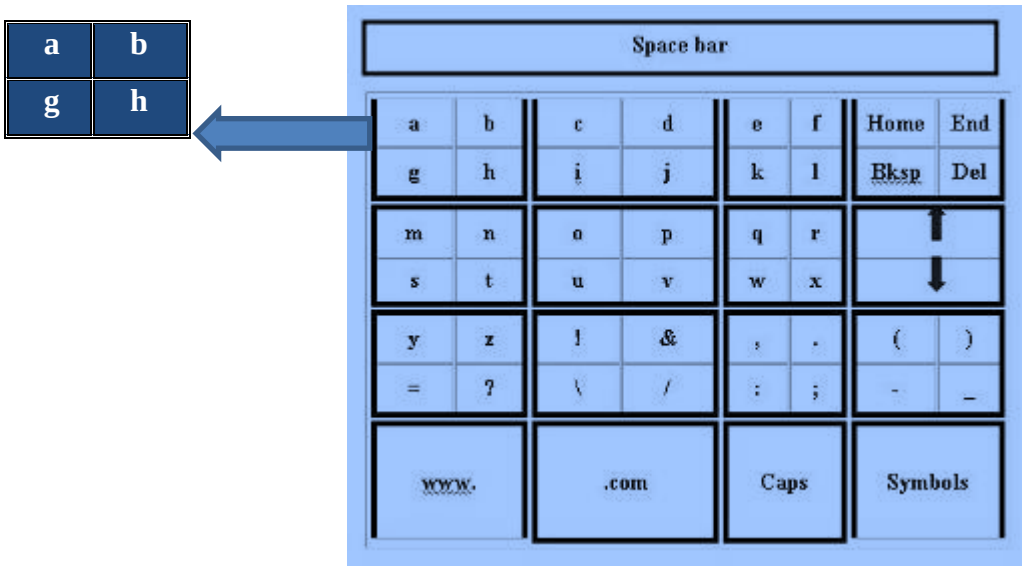


Fig. 6.2: Proposed Keyboard Layout - Lower case letters

As a novel layout, it is proposed that the character set of the on-screen keyboard is organized in blocks with each block containing up to a maximum of 4 characters (Fig. 6.2). Hierarchical navigation and selection method is used across and within the specially organized character blocks in such a way so as to reduce the number of key-strokes required for navigation and selection. The blocks are navigated using the four arrow keys of the remote (up, down, left, right). Once user has navigated to a particular block, they can choose the desired key in the block of four characters using the Up Left, Up Right, Down Left and Down Right keys. For example, one can move from the “o-p-u-v” block to the “c-d-i-j” block by pressing the Up key and then move from “c-d-i-j” block to the “a-b-g-h” block by pressing the Left key. Once on the “a-b-g-h”, block, “a” can be selected by pressing Up Left key, “b” can be selected by pressing the Up Right key, “g” can be chosen by pressing the Down Left key and “h” can be chosen by pressing the Down Right key.

In the proposed layout, the horizontal groups of key-blocks are termed as rows, vertical groups of key blocks are termed as columns and individual characters in a key block are termed as cells. A special meaning can be attached to a cell which, on selection will manifest some special behavior rather than typing the content on that cell.

6.3.1 Proposed Algorithm for Optimal Layout

To design the block-based hierarchical layout and navigation concept introduced in the previous section in an optimal fashion, the problem can be defined as organizing a given numbers of characters or symbols or character-sets optimally so that number of required keystrokes is reduced. The problem is represented mathematically as below.

Total No. of Character Cells = *T*, Total no. of rows of key-blocks = *R*

Total no. of columns of key-blocks = *C*, Total no. of Character Cells in a key-block = 4

T = R x C x 4, Then in the worst case scenario, moving from one character cell to another character cell will take maximum *K = (R+C+1)* keystrokes.

Hence, the desired solution boils down to finding *R* and *C* for which *K* is minimum. Fig. 6.3 provides flow-chart for the proposed algorithm for finding the optimal *R* and *C*.

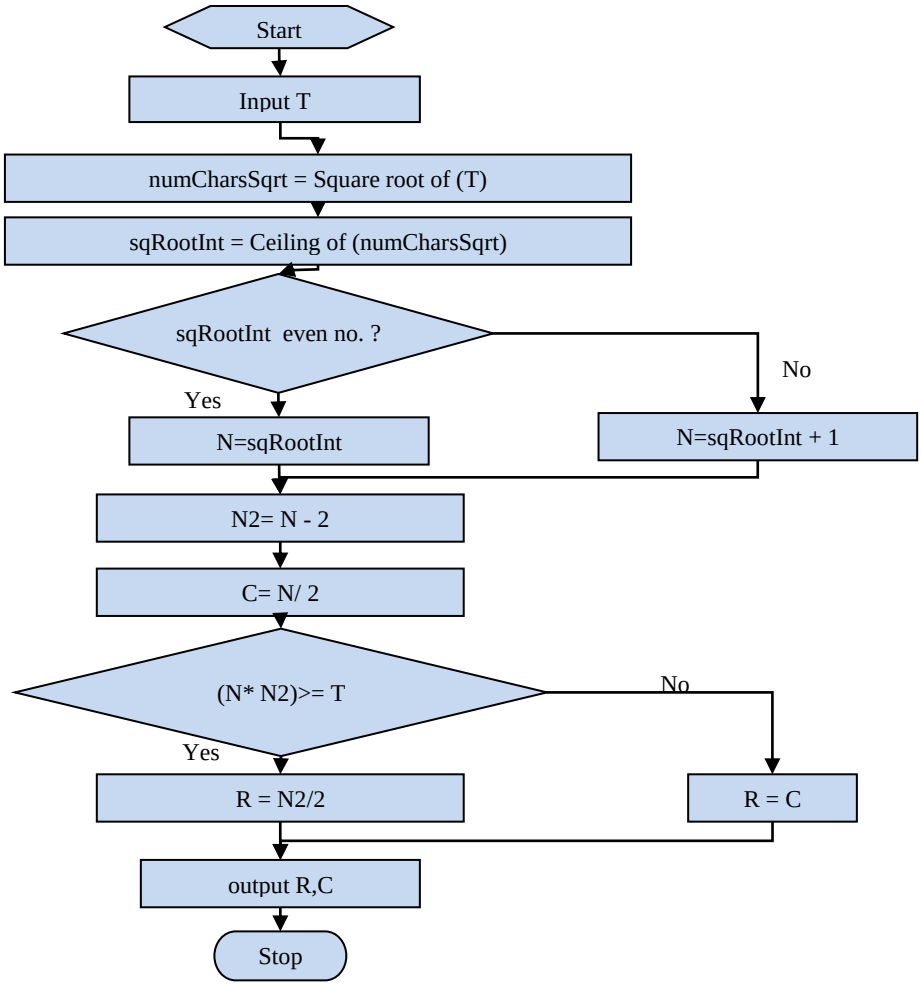


Fig. 6.3: Algorithm Flowchart for Keyboard Layout Decision

It can be deduced from this algorithm that for T number of characters, there exists a square number N^2 having an even number square root and is equal to T or just the next square number after T. The reason for having even number square root is that in both horizontal and vertical directions, there can be 2 cells per key-block.

As example, in a typical "QWERTY" keyboard of laptop or PC, *T=54* (only basic characters) arranged in 4 rows and 14 columns, thereby requiring approx. (*14+4=18*) keystrokes in the worst case to reach from say the most bottom-left character to most up-right character.

On the contrary, for the proposed scheme, for *T=54*, if walking through the algorithm,

sqrtInt = 8, which is an even number and hence *N=8*.

Then, *N2=6*, *C=4*. Since *N*N2=48* is less than *T=54*, *R = N/2 = 4*.
So finally, *R=4* and *C=4*. In such a scenario, one will need maximum (*4+4+1 = 9*) keystrokes to reach from one corner of the keyboard to the other corner (worst case scenario). Hence quite a significant amount of benefit of reduced keystrokes can be obtained in the proposed hierarchical key organization.

For another example of *T=48*, walking through the algorithm,
sqrtInt = 7, which is an odd number and hence *N=8*.
Then, *N2=6*, *C=4*. Since *N*N2=48* is equal to *T=48*, *R = N2/2 = 3*.
So finally, *R=3* and *C=4*. In such a scenario, one will need maximum (*3+4+1 = 8*) keystrokes to reach from one corner of the keyboard to the other corner. This layout was used in HIP.

6.3.2 Implementation Details

The on-screen keyboard was implemented on HIP with 48 basic keys placed in a 3 x 4 layout. There were two variants for character placement – Layout 1 had all consecutive characters in one block (Fig. 6.4, e.g. a-b-c-d in one block) and Layout 2 had all consecutive characters in one row (Fig. 6.5, e.g. a-b-g-h in one block). For conducting the user study, a traditional "QWERTY" layout on-screen keyboard (Fig. 6.6) was also implemented for comparison purpose.

All the three layouts were used in the user study to figure out the best suited layout along with collecting user feedback for improvement. The user study is covered in details in the next section.

Fig. 6.4: Proposed Layout - 1

Fig. 6.5: Proposed Layout – 2

Fig. 6.6: Traditional QWERTY Layout

6.4 User Study

6.4.1 Methodology

Three separate user studies were conducted for evaluating the onscreen keyboard.

USER STUDY 1: BASIC BENCHMARKING STUDY

A user study was conducted to gather user feedback on the proposed layout 1 as compared to the "QWERTY" layout. Multiple techniques were employed to gather feedback from different users with a view to discover the strengths and weaknesses of the proposed design and implementation. It involved two aspects – a) Comparison of the traditional QWERTY on-screen keyboard layout with the proposed layout 1, and b) User Perception Study for the proposed layout for finding out improvement areas.

Users

25 users were selected from different age group having different keyboard exposure, however the scope of the evaluation was kept limited to finding average user response and detailed demography based study was kept out of scope.

Tasks: The users were asked to type particular short message service (SMS) text, Email and uniform resource locator (URL)texts on HIP using a standard QWERTY on-screen keyboard and the proposed on-screen keyboard. Users were asked to type the text "The quick brown fox jumps over a lazy dog" on SMS and Email applications. This particular sentence was used because it contains all the alphabets in English. Users were also asked to type "www.google.com" as the URL in the Internet Browser Application. The time taken to type was noted.

These users were also asked to get familiar with the proposed on-screen keyboard via practice and same data were recorded after they became conversant.

Users were also asked to give some qualitative feedback and suggest improvements to the proposed layout. Based on the feedback, quite a few suggestions were incorporated in the subsequent design. These are elaborated in **Appendix A**, section A.3.

USER STUDY 2: DETAILED STUDY USING KLM-GOMS MODEL

The basic benchmarking study presented above is a good starting point to prove the usefulness of the proposed layout, however it is not formal enough to be repeated across designs and is not accurate enough to compare between two closely placed layouts like Layout-1 and Layout-2, Hence it is proposed to use the KLM-GOMS model [10] for evaluation. The task at hand is designed using basic operators defined in KLM model. A KLM consists of a number of operators. There are five operators which are of relevance here, which are listed in Table 6.1 along with average time taken for the operations as quoted in the literature.

Table 6.1: KLM Operators

Operators	Description	Time in sec
P	Pointing a pointing device	1.10
K	Key or button press	0.20
H	Move from mouse to keyboard and back	0.40
M	Mental preparation and thinking time.	1.35

For current experiments, one can use the values of the operators K and M as-is as the operation remains similar in current case. The definition of the P operator has been modified keeping in mind the present scenario.

In original KLM model, P operator represents pointing with a pointing device at a particular position of the screen, excluding the button press. In the present scenario, however there is no pointing device – instead there is the two key-stroke based block navigation and key selection process (as proposed in section 6.3.1). Hence P has been redefined as the total time taken in finding a key on a particular layout and moving the focus to select the block containing that particular key. To estimate the value of P, a user study was conducted. A group of 20 users of

different age group were selected for the study. They were given three different layouts (layout1, 2 and QWERTY) one at a time. During the study, a tape recorded message consisting of 20 randomly selected alphabets was played. The users were instructed to focus on the particular block containing the alphabets. The time taken to finish the session was noted using a stop watch. To reduce the error as much as possible, the average value was taken for each user. For Layout-1, value of P obtained was 1.77s and for Layout-2, it was 1.73 sec).The value of P for standard QWERTY keyboard layout has been adopted as 1.10 sec as defined in original KLM model.

The H operator was found not to be useful for an on-screen layout using remote control as there is no concept of mouse here. A new parameter F was introduced for measuring the time required for finger movement. The justification of including this parameter has been explained in [14]. The average value of H was found to be 0.22sec for all layouts.

The evaluation served as a basic test to see whether KLM-GOMS can be applied for onscreen keyboards. Once its applicability is established, this concept can applied to evaluate the on- screen layouts for an email sending application. Gmail was taken as the mail server using the browser on HIP and task was to send an email with a text “Hi”, to a new email address.

The KLM-GOMS for sending an email is shown in Table 6.2.

Table 6.2 : KLM-GOMS for Email Sending Application

Goal	Subgoals
Compose and send an email	Open browser
	Open gmail server & login
	Compose mail
	Dispatch

During the process of calculation and evaluation, some keyboard level operations were defined. These are actually the basic operations to be performed while using the layouts (1 and 2) with the remote control and thus served as the basic building blocks of the GOMS model. These can be looked upon as a composition of different KLM operators explained in Table 6.1. The values as computed from the KLM-GOMS for different layouts as given in Table 6.3.

Table 6.3: Basic Operations for KLM-GOMS

Operations	Time for Layout-1 in sec	Time for Layout-2 in sec	Time for QWERTY in sec
Open/close onscreen keyboard layout.	0.4	0.4	0.4
Find any key	1.07	1.03	1.1
Move focus to select a key	0.7	0.7	2.6
Move finger to the corner keys	0.2	0.2	0.2
Enter a character using keyboard	2.17	2.13	4.0

KLM-GOMS gives a scientific model for predicting the user experience during the on-screen keyboard operation. User studies were then conducted for all the three different on-screen layout under consideration with the basic operations for KLM-GOMS being the guiding factor for measurement. The results are outlined in the next section along with a comparison with the results predicted from KLM-GOMS.

6.4.2 Results and discussion

USER STUDY 1

Here, the proposed novel onscreen keyboard layout (Layout-1) is compared against the traditional QWERTY layout. The measurements were done in terms of characters per minute (CPM). The total number of users taken was 25. The CPM values were calculated for each user and were

plotted for both standard QWERTY onscreen keyboard and for Layout-1. The results are shown in Fig. 6.7, from which it can be observed that there are significant improvements in CPM for the proposed layout over the QWERTY one indicating that the proposed layout is better. After doing some outlier rejection, the average percentage improvement using the proposed keyboard over QWERTY on-screen keyboard was found to be 27% for Email, 41% for SMS, and 38% for URL typing.

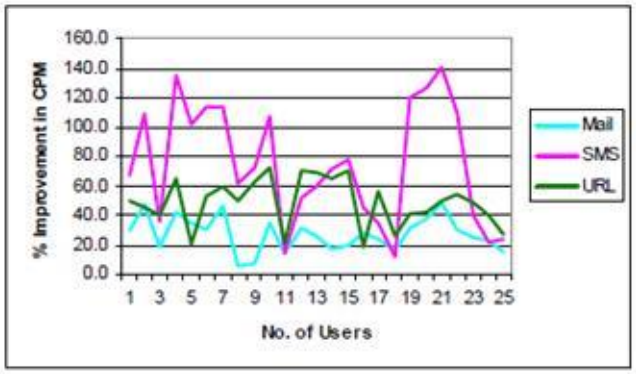


Fig. 6.7: Improvement of Layout-1 over on-screen QWERTY Layout

Another study was conducted to understand the effect of familiarity and practice on the user experience of Layout-1. Fig. 6.8 shows the results after some practice compared to the initial runs. It shows an average improvement of 31.8% for email, 29.9% for SMS and 16.69% for URL typing.

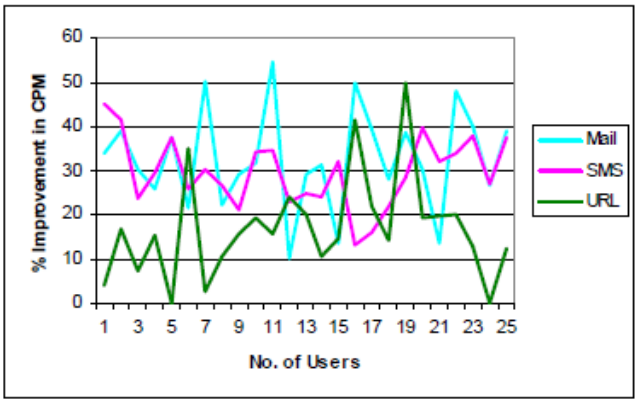


Fig. 6.8: Improvement in Layout-1 after practice

From the results of User Study 1, it can be established that the proposed Layout-1 indeed provide a significantly improved text entry speed compared to traditional QWERTY layout. It is also quite evident that since the proposed layout is initially unfamiliar to the users, the user experience also improves over time with more practice.

USER STUDY 2

For comparing normal text entry through different layouts in a more scientific way using KLM-GOMS, a separate user study was conducted as per the methodology outlined in section 6.4.1. Six phrase sets were selected randomly from MacKenzie's test phrase set [15]. Users were given an initial familiarization phrase and then asked to enter six phrases at one go. Time taken by each user and the number of keystrokes required to type the phrase were recorded. A total of 20 users were considered for the study.

For normal text entry, the user study reveals that the Layouts-1 and layout-2 allow much faster text entry compared to onscreen QWERTY layout. The percentage improvement in time for the two proposed layouts over traditional onscreen QWERTY layout was more than 40% (Table 6.4). The percentage improvement for Layout-1 over Layout-2 was found to be approximately 2%. The

results also show significant agreement with the theoretical results predicted by KLM-GOMS model.

Table 6.4: User Study Results for Normal Text Entry

Layouts	% improvement in typing speed from Experiment	% improvement in typing speed as predicted from KLM-GOMS
Layout 1 over QWERTY	44.23	45.75
Layout 2 over QWERTY	45	46.75
Layout 2 over layout1	2	1.84

To validate the values obtained from KLM-GOMS model for the email task, the scenario of sending an email is modeled with proposed parameters and then compared with the actual values obtained through user study. A group of 20 users of different age groups and different computing background were selected. They were said to send an email using conventional QWERTY on-screen keyboard, as well as using two layouts mentioned above. Their feedback was captured, recorded, analyzed and used to evaluate the three design layouts.

For calculating the time taken for each layout, the following assumptions are made:

- Length of website address = 9 characters
- Length of login id = 6 characters
- Password = 8 characters
- Recipient's address = 15 characters
- Length of subject of email = 15 characters

The time taken for different layouts is given in Table 6.5 below.

Table 6.5: User Study Results for Email Application – Time Measurement

Sub-goals	Time for layout 1 in sec	Time for layout 2 in sec	Time for QWERTY in sec
Open browser	0.5	0.5	0.5
Open gmail server & login	45.3	44.5	82.1
Compose mail	68.03	65.87	115.1
Dispatch	0.4	0.4	0.4

These measurements were then combined and compared with those predicted by KLM-GOMS model. As seen from Table 6.6, the study results map closely to the KLM-GOMS results. Here also, like the text entry task, the two proposed layouts exhibit close to 40% improvement over the traditional QWERTY layout. Additionally, as was found in the text entry user study, here also Layout-2 scored over Laout-1.

Table 6.6: User Study Results for Email Application – Comparison of Layouts

Layouts	% improvement in typing speed from Experiment	% improvement in typing speed as predicted from KLM-GOMS
Layout1 over QWERTY	42.2	35.23
Layout2 over QWERTY	43.4	37.2
Layout 2 over Layout1	3.18	2

From the results, it is quite clear that the proposed layouts indeed provide a better user experience compared to traditional QWERTY layouts. It is also quite evident that Layout-2 is a slightly better option than Layout-1. This result also validates our proposed extension on P and F parameters of KLM-GOMS.

6.5 Conclusion

In this chapter a novel hierarchical navigation based on-screen keyboard layout especially suited for operation with infra-red remote controls is presented in the context of HIP. It addresses the user experience concern for text entry that was found during the HIP user study (**Chapter 2**). As the main contribution, a mathematical algorithm to arrive at the optimal layout for such on-screen keyboards was presented first. The proposed layout was subjected to extensive user study and compared with traditional layouts to show that the proposed layout indeed results in faster text entry. In addition to

direct measurement of text entry time, the user study also adopted the much-researched KLM-GOMS model as the theoretical basis of the study. As another contribution from our work, the KLM-GOMS was extended (P and F parameters) for hierarchical on-screen keyboard layouts. The user results showed good agreement with the KLM-GOMS model. It not only validated our proposed extensions on KLM-GOMS, but also formally showed that the proposed layout provides much better user experience compared to traditional "QWERTY" layouts.

References

[1]. Konstantinos Chorianopoulos, George Lekakos, Diomidis Spinelli, "Intelligent User Interfaces in the Living Room: Usability Design for Personalized Television Applications", *IUI'03,* Jan 2003, Miami, Florida, USA.ACM 1-58113-586-6/03/0001

[2]. Tiiu Koskela, Kaisa Va¨a¨na¨nen, Vainio-Mattila, "Evolution towards smart home environments: empirical evaluation of three user interfaces", *Pers Ubiquit Comput* (2004) 8, Springer-Verlag, June 2004

[3]. Simon, "On-screen keyboard," US Patent No. 20080303793 A1, *by Microsoft Corp.,* June 2007

[4]. Daniel, W. Steven, "Intelligent default selection in an on-screen keyboard", US Patent No. 7130846 B2, *by Microsoft Corp.,* Dec 2004

[5]. G. Kevin J, Z. Thomas J, "On-screen remote control of a television receiver", US Patent No. 5589893 A, *by Zenith Electronics Corp.,* Dec. 1994

[6]. Boyden David, "On screen display for alpha-numeric input", US Patent No. 7716603, *By Sony Corp., Dec.* 2005

[7]. Geleijnse, G., Aliakseyeu, D., and Sarroukh, E, "Comparing text entry methods for interactive television applications", *Proceedings of the Seventh European Conference on European interactive Television Conference*, Leuven, Belgium, June 2009, *EuroITV '09*. ACM, New York, NY, 145-148

[8]. Iatrino and S. Modeo, "Text editing in digital terrestrial television: a comparison of three interfaces", *Proceedings of EuroITV'06,* Athens, Greece,2006

[9]. Ingmarsson, M., Dinka, D., Zhai, S., "TNT-A numeric keypad based text input method", *Proc. CHI 2004: ACM Conference on Human Factors in Computing Systems.* Vienna, Austria. *CHI Letters 6(1),* 639-646,ACM Press.

[10]. Card, S. K., Moran, T. P., Newell, A., "The Keystroke-Level Model for User Performance Time with Interactive Systems". *Comm. ACM 23, 7*. 396-410. 1980.

[11]. Bälter, O., "Keystroke Level Analysis of Email Message Organization", *Proc. CHI'00*. ACM Press. 105-112. 2000.

[12]. Manes, D., Green, P., Hunter, D., "Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models", *Technical Report UMTRI-96-37*. The University of Michigan Transportation. 1996.

[13]. Teo, L., John, B. E., "Comparisons of Keystroke-Level Model Predictions to Observed Data", In *Extended Abstracts CHI'06*. ACM Press. 1421-1426. 2006.

[14]. Holleis, P., Otto, F., Hussmann, H., and Schmidt, A. (2007). "Keystroke-level model for advanced mobile phone interaction," in CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 1505–1514, New York, NY, USA.

[15]. MacKenzie, I. S., and Soukoreff, R. W. "Phrase sets for evaluating text entry techniques," in CHI '03 extended abstracts, Lauderdale, USA, 754-755, 2003.

7
Conclusion and Future Work

The main conclusions of this work and the possible areas of future investigation are discussed in this chapter.

7.1 Conclusion

In this thesis, Television has been introduced as a ubiquitous device for masses for information access, suited for developing countries like India. However, the emerging markets like India are characterized by some unique challenges like low bandwidth / low Quality-of-Service (QoS) of the available wireless networks, extreme cost-consciousness of the users and lack of computer literacy among masses. Most of the available solutions for Connected TV are targeted towards high-end markets and none of them really address these challenges posed by the developing markets. The main motivation of this thesis has been to create a connected television solution for developing markets addressing these challenges.

The high-level challenges like low bandwidth / low Quality-of-Service (QoS) of the available wireless networks, extreme cost-consciousness of the users and lack of computer literacy among masses translate into some unique technology challenges as listed below –

1. Lack of open source software middleware and framework on low-cost processors families
2. Needs for Video Chat Application to work with user-acceptable quality even when the available bandwidth and QoS of the Wireless Access network fluctuates to a low level.
3. Need for computationally efficient access control and DRM solutions in the Video Chat and Video Content Sharing, especially for medical and education applications.
4. Need for the browsing experience to seamlessly blend into the broadcast TV viewing experience, thereby requiring intelligent mash-ups of broadcast TV and internet content.
5. Need for an easy to use on-screen keyboard for text entry on TV using remote control.

In **Chapter 2**, a novel low-cost internet-enabled TV application platform called Home Infotainment Platform (HIP) is introduced. The proposed solution provides a low-cost information access device using a low-cost ARM CPU based platform with application focus on infotainment, healthcare and education. HIP is already deployed in pilot scale in the Indian and Philippines' market. The engineering contributions on HIP are elaborated in **Appendix A**.

As the main scientific contribution here, a flexible and scalable multimedia framework on top of the low-cost ARM-based CPU of HIP is proposed for quickly developing variety of applications,

thereby keeping the overall development cost low. As a further contribution, a user study in the field was conducted with HIP and the results analyzed. Based on this analysis, the challenges and needs outlined in **Chapter 1** are ratified. Some of these challenges are addressed in subsequent chapters.

In **Chapter 3**, the problem of low-QoS cellular networks in India for Video Chat is addressed. A multi stage adaptive rate control over heterogeneous network for a H.264 based video chat solution is proposed. As the main contribution in this chapter, three stages of design enhancements and methodologies are proposed –

a) Probe-packet-pair delay based sensing of network condition

b) Adaptive rate control of audio

c) Video compression and adaptive packet fragmentation for video and audio packets.

The novelty in each these respective modules are –

a) An experimental heuristics based mapping of effective bandwidth to probe packet delay

b) An efficient video rate control using automatic switching between frame and macro-blocks

c) An adaptive scheme for audio/video packet fragmentation.

The system is tested using HIP in real network scenarios of 2G modems and ADSL. Results supporting the novelty claims of all the three sub-systems is also presented and analyzed.

In **chapter 4**, a set of novel watermarking and encryption algorithms are presented that can be used for DRM and access control of security-sensitive video-centric applications of HIP like multimedia content distribution in distance education and patient-doctor video chat in remote medical consultation. The main differentiating feature of both the proposed watermarking and encryption algorithms are their low-computational complexity without compromising on the video quality, while still providing adequate security.

For watermarking, in addition to proposing a novel watermark embedding algorithm, other algorithms required for a complete system like frame level integrity check after embedding the watermark, finding space and location inside the video for embedding the watermark, handling images and text strings separately etc. were also designed and implemented. An implementation of watermarking evaluation tool and a novel methodology to evaluate watermark against attacks is also presented. Finally, a detailed set of experimental results are presented using the tool and methodology introduced to prove the low computational complexity, preservation of video quality and security robustness of the proposed watermarking algorithm.

For encryption, a novel two-stage algorithm is proposed that keeps the computational overhead low through doing separate header encryption and reusing the flexible macro-block reordering (FMO) feature of the H.264/AVC as the encryption operator. Security analysis of the algorithm is presented to prove their robustness against attacks. It is also shown through experiments that in spite of being computationally efficient and secure, the proposed algorithm does not deteriorate the video quality.

In **Chapter 5**, the problem of user dissatisfaction of using Internet on TV is addressed. Here a novel concept of mashing up of related data from internet by understanding the broadcast video context is presented along with three types of applications on Connected TV. As the main contribution, three novel, low-computational complexity methodologies for identifying TV video context are proposed –

a) Channel Identification using logo recognition and using it for an EPG application form the web for analog TV channels

b) Detecting text in static screens of Satellite DTH TV active pages and using it for an automated mode of interactivity for the end user

c) Detecting text in form of breaking news in news TV channels and using it for mashing up relevant news from the web on TV.

Experimental results show that the applications are functional and work with acceptable accuracy in spite of being computationally efficient. For developing nations this can be the way to bring power of internet to masses, as the broadcast TV medium is still primarily analog and the PC penetration is very poor. The proposed solution is definitely one way to improve the poor internet experience reported in the HIP user study in **Chapter 2**.

In **Chapter 6**, the problem of difficulty in entering text on TV using a remote control is addressed. Here a novel hierarchical navigation based on-screen keyboard layout especially suited for operation with infra-red remote controls for HIP is presented. As the main contribution here, a mathematical algorithm to arrive at the optimal layout for such on-screen keyboards was proposed. The proposed layout was subjected to qualitative and quantitative user studies and compared with traditional layouts to show that the proposed layout indeed results in faster text entry.

Additionally, as a formal methodology, the user study also adopted the much-researched KLM-GOMS model as the theoretical basis of the study. As another contribution from our work, the KLM-GOMS was extended for hierarchical on-screen keyboard layouts. The results showed that the proposed layout provides much better user experience compared to traditional "QWERTY" layouts. The user results showed good agreement with the values predicted by the KLM-GOMS model, thereby proving the efficacy of the proposed KLM-GOMS extensions.

The user study also helped in improving the layout further and providing a few engineering enhancements on the proposed on-screen keyboard, which are detailed in Appendix A.

7.2 Future Work

This is a new area of technology application with actual pilot deployments under way. The feedback from the pilot deployments are going to shape the requirement for the future. As a natural extension of the current work based on initial feedbacks received from pilot deployments, the following areas can be looked into for future work.

On the engineering front, there is need to bring down the cost of the box further through hardware optimization and intelligent hardware-software integration without compromising on performance. Recent advancements in mobile processors and mobile operating systems give us scope to improve in this direction and work has already started on this area.

On the scientific front, even though a few of challenges have been solved and presented in the thesis, there is scope for additional work in this area, which can be taken up in near future. These are listed below.

There is scope for using better Quality of Experience (QoE) measures instead of PSNR used in **Chapter 3** for improving the adaptive rate control performance which in turn can improve the video chat experience under poor network conditions. Qualitative research methods on end-user experience can be used for measuring the QoE through user studies. The network sensing module as presented here can also be enhanced further by monitoring the statistics of the received TCP/UDP data packet thus exploiting the cross layer architecture fully. The adaptive rate control can also be extended to the multipoint multicast scenario, which would be the case for multi-party video conference.

In future, more work can be done on further improving the robustness of the watermarking algorithm and encryption algorithm in **Chapter 4** against newer kind of attacks. There is scope for extending the security analysis of the encryption algorithm beyond brute-force attacks. The area of audio watermarking as content DRM mechanism can also be explored.

The Internet-TV mash-up applications in **Chapter 5** can be further extended to second-screens (mobile, tablets), where information from internet can be rendered on mobile / tablet screens automatically based on the current program being watched on the TV. There is also requirement for seamless transcoding and transfer of streaming content from TV to mobile/tablets and vice versa. Integrating social media into broadcast TV is another upcoming area to look at. On the design front, there is scope for further reducing the computational complexity of the context detection algorithms through better design of algorithms, efficient usage of the hardware accelerators present in the platform and innovative software-hardware partitioning. There is also scope for user study based

qualitative research for proving the improvement of end-user Internet experience through the proposed solutions.

There are other ways to improve the user interface as an extension to the work presented in **Chapter 6**. One approach is to include a predictive keyboard in the on-screen keyboard design and back it up with adequate user studies. Other approaches may include voice interfaces and gesture controls. Since these kinds of interfaces are already becoming available on mobile phones, it will be interesting to explore using mobile phones as the remote control for TV.

On the generic engineering design of the platform, it was found that significant challenges exist in balancing the contradicting triad of requirements in form of cost, feature and performance, especially in the face of ever improving hardware functionalities and changing market dynamics. Based on the experience and knowledge gathered in the HIP project, research work can be conceptualized on a scientific methodology for choosing the optimal configuration for hardware and software keeping the cost, feature and performance requirements in balance.

Appendix A
Home Infotainment Platform

A.1 System Description

A Home Infotainment Platform (HIP) has been developed that provides consumers access to interactive services over the television at a reasonable cost using standard Internet connection through 2G/3G wireless and ADSL. It is an information and communication platform that provides consumers with basic computer functionalities (information access, entertainment and collaboration) on a television set. The proposed solution prioritizes the requirements of a non-technology savvy computer user on one hand and the price conscious user on the other. It uses an over-the-top (OTT) box, where the existing TV broadcast is routed through the device via composite A/V in and output is given to the television via composite A/V out. The four USB ports in the box support devices such as keyboard, mouse, webcam, modem and flash drive. Headphones with microphone can be connected through sockets provided. Fig. A.1 gives the basic overview of the platform. The platform uses Linux as operating system and other open source components to keep the box cost low.

The solution supports core applications like Internet Browsing, Photo Viewer, Music Player, Video Player, SMS and Video Chat. It also supports socially value-adding applications like remote medical consultation and distance education. All applications are developed on top of the novel framework proposed in Chapter 2 and are presented in detail in section A.2.

A.1.1 Hardware Details

Processor family, RAM and Flash size are chosen keeping the application feature requirement in mind while minimizing the cost/performance ratio. Various systems were studied and prototype reference designs were evaluated before arriving at the final configuration. The interfaces are decided based on application requirements. After careful analysis of all the requirements, cost and availability, the final hardware configuration was chosen as below –

Processor - Texas Instrument's DaVinci DM 6446 with dual core (300 MHz ARM 9 with 600 MHz 64x DSP) – upgraded to 720 MHz ARM-CortexA8 in subsequent version
RAM – 256 MB
Flash – 128 MB
USB 2.0 ports – 4 (USB modem, USB Flash Drive, USB Webcam and optional Wireless USB Keyboard/mouse dongle)
100 Mbps Ethernet – 1 (optional port used for ADSL broadband, if available)
Interfaces – A/V in/out, Microphone in / Headphone out, IR in, VGA out (optional)

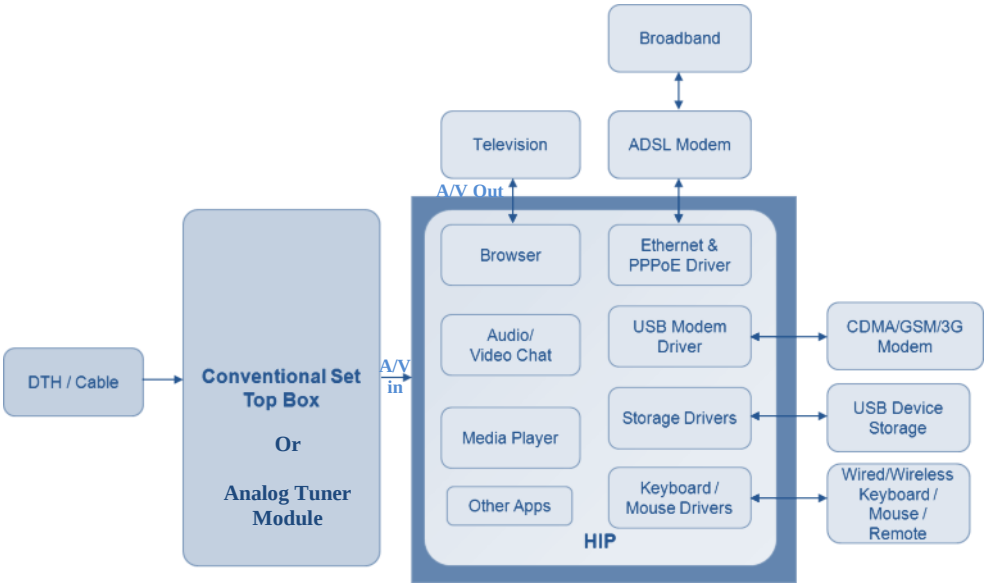


Fig. A.1: Home Infotainment Platform System Block Diagram

A.2 Applications

The basic applications on HIP consist of basic infotainment applications like Internet Browsing, Media Player, SMS on TV, Video Chat and socially value-adding applications like remote medical consultation and distance education. All the applications are developed on top of the framework proposed in Chapter 2 as per the mapping given in Tables 2.1, 2.2 and 2.3. The features of the applications are outlined below.

A.2.1 Browser

- Flash Lite 3.0 with Action Script 2.0
- HTML 4.0 with Java script and ECMA script support / CSS 1.0
- Native Java Script extension support
- File explorer / Simultaneous TV and Browser /Tabbed browsing / Proxy Support
- Navigation through remote control arrow keys with fit-to-width rendering

Fig. A.2 gives example screenshots of invoking browser from the main menu, blended TV and browser application.

A.2.2 Media Player

- Audio - MP3, AAC, FLAC, OGG-Vorbis
- Video - MPEG1 video (VCD), MPEG4, H.263, H.264
- Image - JPEG, GIF, PNG, BMP
- Auto-listing of content based on media type and remote control based navigation

Fig. A.3 gives example screenshots of Picture Viewer, Music Player and Video Player on HIP.

Fig. A.2: Main Menu and Browser on HIP

Fig. A.3: Picture Viewer, Music Player and Video Player on HIP

A.2.3 Video Chat

- Peer-to-Peer connection over UDP
- Supports CIF and QCIF resolution
- Simple mobile no. based connection setup

Fig. A.4 gives example screenshots of the video chat application (session setup and live session).

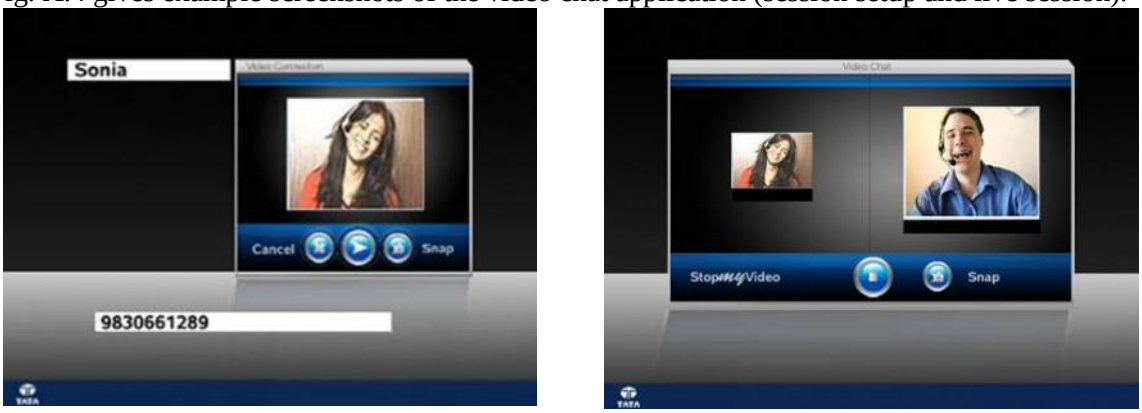


Fig. A.4: Video Chat on HIP

A.2.4 SMS on TV

- CDMA/ GSM support
- Inbox and Address Book support from SIM
- Option of SMS sending while watching TV through alpha blending

Fig. A.5 gives the example screenshots of the SMS application on HIP.

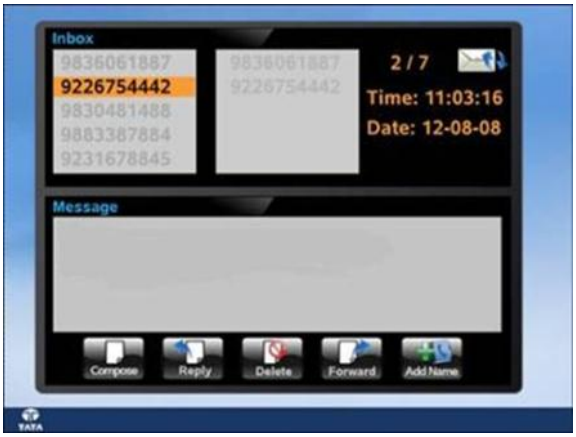


Fig. A.5: SMS on HIP

A.2.5 Remote medical Consultation

Telemedicine today is viewed as a cost-effective and convenient alternative to face-to-face doctor consultations. In developing countries, many people do not have access to expert and specialist medical care. These people may have access to general physician or other health professionals who cannot successfully treat and cure some of their ailments and in those cases expert or specialist doctors' advice is required. In such cases, specialist doctors can use the tele-consultation facility to remotely communicate with a patient and his local physician or the medical technician, remotely look at relevant pathological / clinical / physiological data and video of the patient, and advice a prescription. To enable this successfully, the relevant physiological and pathological data and image from different sensors including a camera and medical instruments need to be transferred from the remote patient's local facility to an expert doctor.

A variant of HIP is used to provide such a solution. Fig. A.6 gives the overview of the proposed system architecture. As seen from the figure, patient data can be uploaded to the server via internet. The data can either come from a medical instrument like ECG / Digital Stethoscope / Blood pressure Monitor / Pulse Oximeter connected to HIP over USB / Bluetooth over USB / Wireless RF over USB, or from diagnostic reports stored in a USB storage device. The expert doctor can access the uploaded patient data from his/her PC / Laptop Browser. If required the expert doctor can initiate a Video Chat session with the patient.

During the design and implementation of this application, one requirement that came up was that of securing the patient-doctor interaction over video chat due to confidentiality issues. This in turn translates into a need to implement encryption and watermarking with low computational complexity on HIP. This problem is addressed in **Chapter 4**. Performing video chat over low-QoS networks also is a challenge for remote medical consultation which is addressed in **Chapter 3**. A few screenshots of the application is given in Fig. A.7, Fig. A.8, Fig. A.9 and Fig. A.10.

A.2.6 Distance Education

The demand for distance education in developing countries is increasing which is mainly due to the paucity of teachers in rural areas. Despite the fact that 7 million people are engaged in the education system in India, which comprises of over 210 million students in 1.4 million schools, there are severe shortages of skilled teachers in rural areas, for which distance education seems to be a possible remedy. One major challenge in distance education lies in sharing the tutorial (video-audio and associated question-answer (QA)) with individual students or classrooms in the rural schools or homes. The most popular methods used for distance education is based on either the broadband internet protocol (IP) connections or satellite transmission. But internet connectivity in rural India is quite poor and satellite solutions lack interactivity due to absence of return path. Hence a hybrid

approach for distance education solution is taken based on the existing television broadcast network that uses satellite broadcast to send the bandwidth-heavy multimedia content and internet for low-bandwidth real-time interactions (Fig. A.11 and Fig. A.12). HIP is used as the platform for the Student's Station depicted in Fig. A.12.

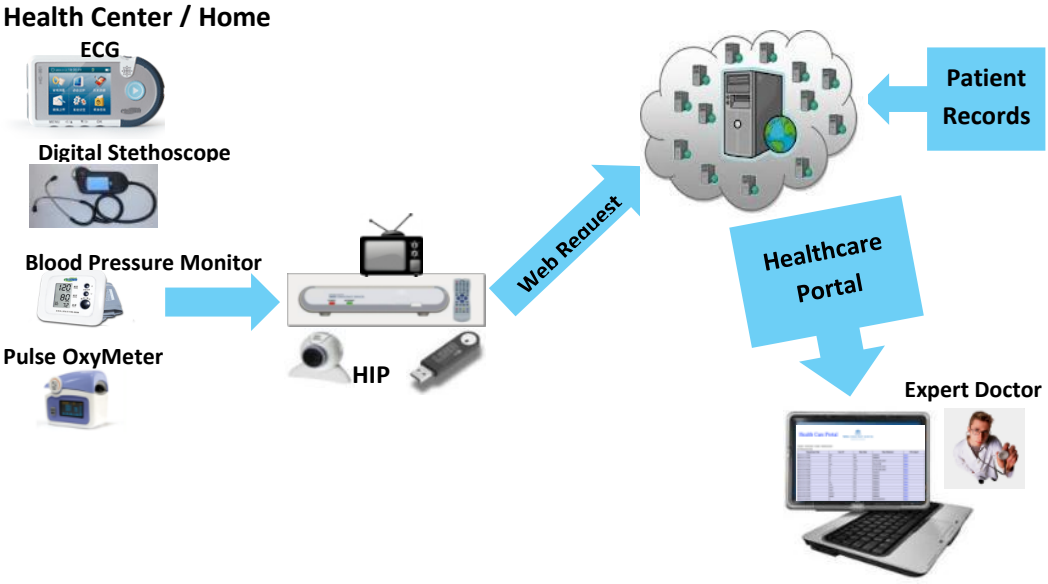


Fig. A.6: Proposed System Architecture

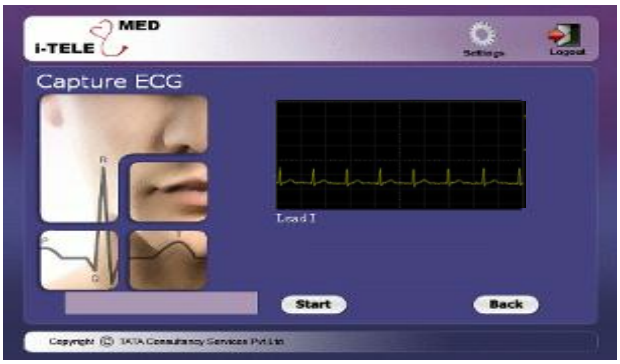


Fig. A.7: Data Capture Screen

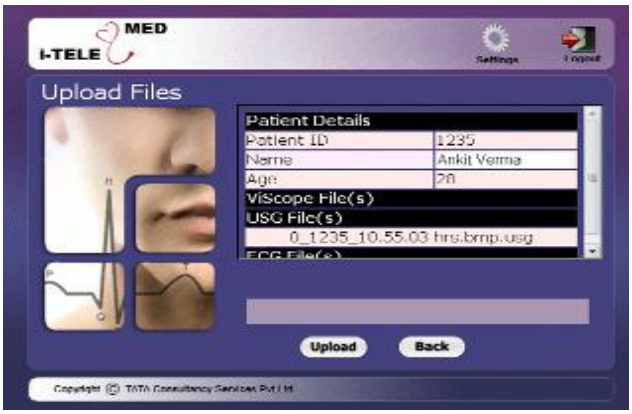


Fig. A.8: Data Upload Screen

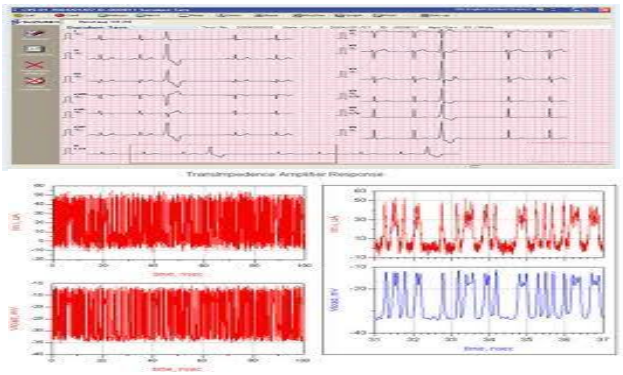


Fig. A.9: Expert Doctor View

Fig. A.10: Video Chat Session

During the design and implementation of this application, one requirement that came up was that of securing the video content delivery handling the copyright and access control issues. This in turn translates into a need to implement decryption and watermarking with low computational complexity on HIP. This problem is addressed in **Chapter 4**.

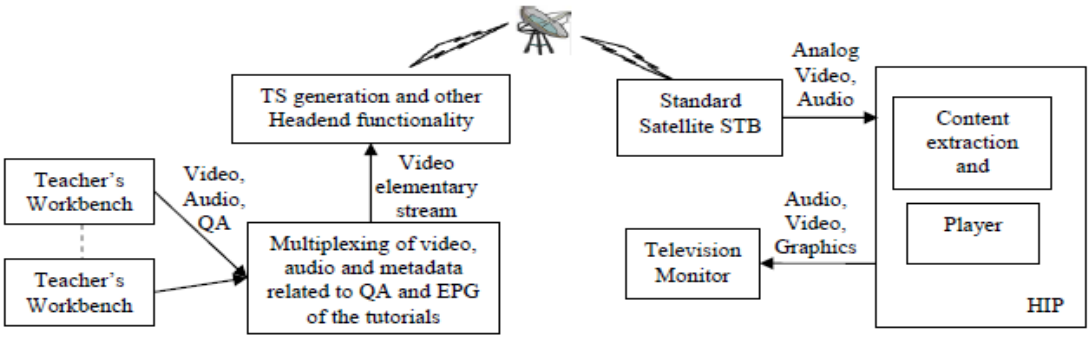


Fig. A.11: Solution Architecture – Satellite Broadcast

A few screenshots of the HIP client application for the video lectures is given in Fig. A.13, Fig. A.14, and Fig. A.15.

A.3 Improvements in HIP on-screen keyboard

For text entry on TV screen, HIP uses a novel on-screen keyboard layout customized for using with remote controls. The keyboard is implemented using JavaScript and is accessible from any application in HIP on a Hot Key trigger on remote. The novel layout of the keyboard is described in detail is **Chapter 6**. Initial feedback on usage of the on-screen keyboard from users of HIP resulted in few improvements in the design and implementation of the on-screen keyboard. These improvements are described below.

1. Providing provision for having fewer than 4 characters, or symbols or character-sets in a key-block for providing greater prominence to some characters, or symbols or character-sets and thus better ease-of-use. Up / down arrow keys and space bar was found to be frequently used keys that can benefit from this feature. Each of Up/down arrow keys was assigned to two cells and the space bar was a complete row of 4 key blocks (Fig, A.16).
2. To reduce the typing effort even further, certain commonly used character sequences like "www." or ".com" is placed together in place of a single character, so that selecting the sequence will type the whole sequence (Fig. A.16).
3. To new special layouts for Capital letters and Symbols were created. These layouts are given in Fig. A.16 and Fig. A.17. Hotkeys for invoking these special layouts were also provided.

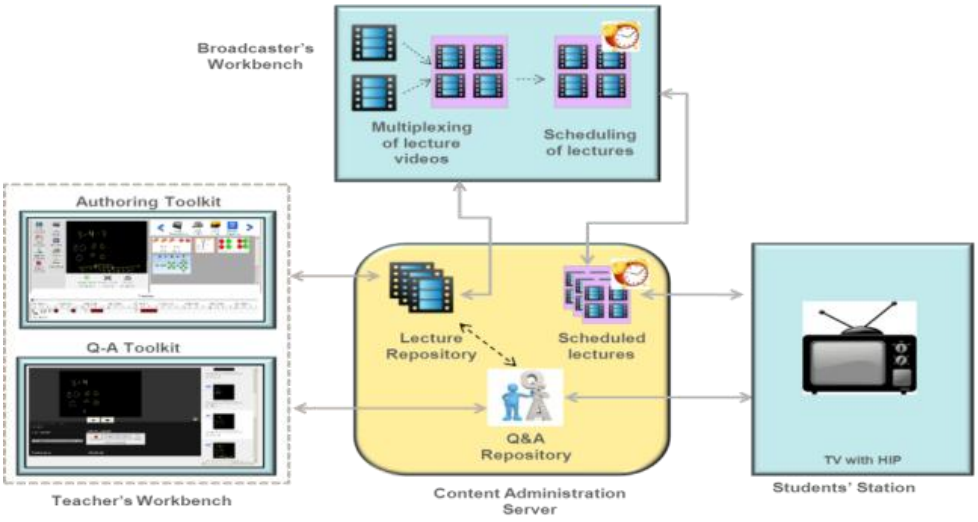


Fig. A.12: Solution Architecture – Internet based Interactivity

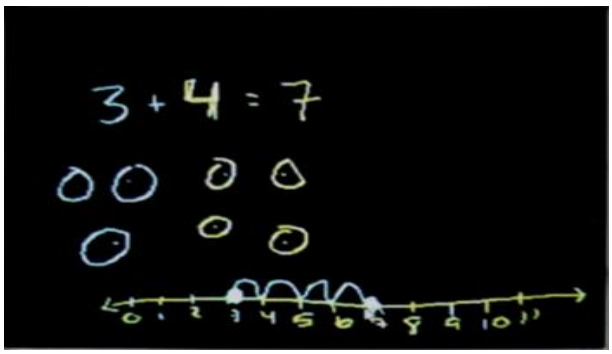


Fig. A.13: Lecture Video

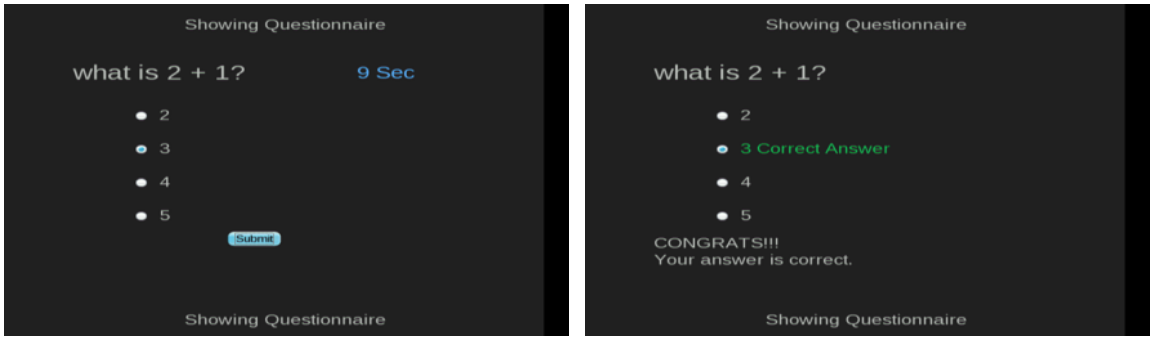


Fig. A.14: Rhetoric Questions and Answer

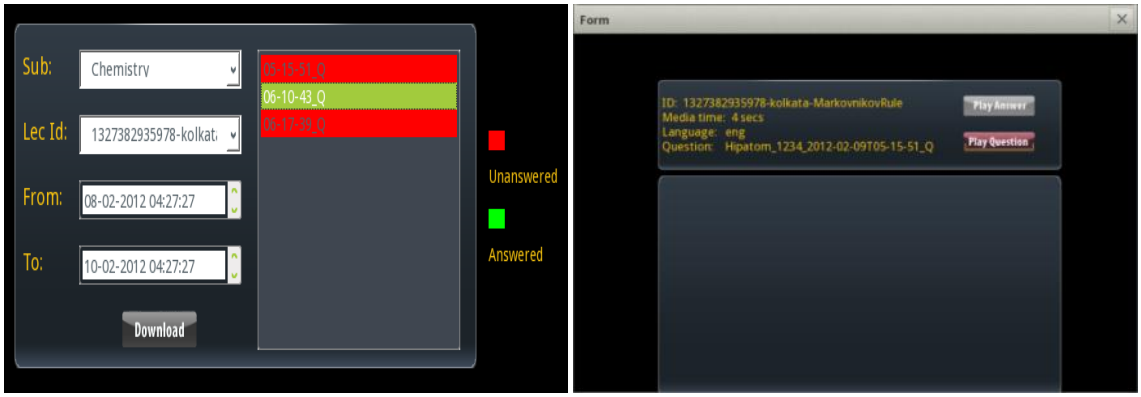


Fig. A.15: Audio Question/Answer Recording and Playback

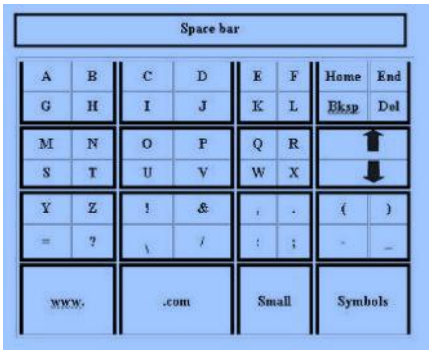


Fig. A.16: Proposed Keyboard Layout - Upper case letters

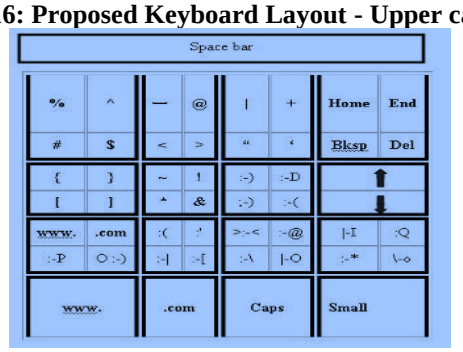


Fig. A.17: Proposed Keyboard Layout - Symbols

Finally, based on the feedback obtained from qualitative user study on HIP (section 6.4.1), some more areas for improvement were found and they were adopted in the keyboard layout of HIP. They are outlined as below.

1. **Most frequently used Smileys:** Initial on-screen keyboard had a number of smileys in the Symbol page which were selected mainly by the developers. Later on the most frequently used smileys were updated to accommodate users' preferences. The final layout design for smileys after accommodating user feedback is shown in Fig. A.18.

2. **The missing @:** In the very beginning the symbol '@' was not available in the frequently used symbols in the main layout. During the user study, users complained about this issue while typing emails. Based on that the symbol '@' was included in the frequently used symbols in the main layout.

3. **Hot keys to speed up typing:** Number of users suggested giving some additional options as hot keys so that some of the most commonly used requirements can be expedited. Accordingly different users' preferences for hot keys were noted and analysed along with the constraint of real estate space on the accompanying remote control. 4 colored keys on the remote were decided to

be used as hot keys – green for typing “space”, red for typing “delete”, yellow and blue to bring up the capital/small letter screen or the symbol screen based on the keyboard context at a particular instance on the remote control.

Fig. A.18: Updated Symbol and Smiley Layout

4. **Visual effect of key press:** During the study, users pointed out that they were not sure or did not know which key has been indeed selected and typed. This at times led user to type a desired letter multiple number of times. As a consequence, it was felt that user should get some kind of visual feedback on the selected character. Accordingly the effect of key press was introduced. As shown in Fig. A.19, when the letter “A” is selected, its appearance momentarily changes to highlight it distinctly.

Fig. A.19: Visual effect of key-press and colored hot keys

5. **Need for co-located help:** The user study result of improving user experience after adequate practicing (section 6.4.2) points to the fact that initially the users have difficulty in figuring out how to use the proposed on-screen keyboard. So as an aid to the users, a brief and crisp pictorial help was embedded along with the on-screen keyboard informing users about how to use it. This is shown in Fig. A.20.

Fig. A.20: Co-located help

Appendix B
List of Publications and Patents

Publications
Conferences

Chapter 2

[1]. Arpan Pal, M. Prashant, Avik Ghose, Chirabrata Bhaumik, "Home Infotainment Platform – A Ubiquitous Access Device for Masses", *Proceedings on Ubiquitous Computing and Multimedia Applications (UCMA),* Miyazaki, Japan, March 2010.

Chapter 3

[2]. Dhiman Chattopadhyay, Aniruddha Sinha, T. Chattopadhyay, Arpan Pal, "Adaptive Rate Control for H.264 Based Video Conferencing Over a Low Bandwidth Wired and Wireless Channel", *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB),* Bilbao, Spain, May 2009.

Chapter 4

[3]. Arpan Pal and T. Chattopadhyay, "A Novel, Low-Complexity Video Watermarking Scheme for H.264", *Texas Instruments Developers Conference,* Dallas, Texas, March 2007.

[4]. T. Chattopadhyay and Arpan Pal, "Two fold video encryption technique applicable to H.264 AVC", *IEEE International Advance Computing Conference (IACC),* Patiala, India, March 2009.

Chapter 5

[5]. T. Chattopadhyay, Aniruddha Sinha, Arpan Pal, Debabrata Pradhan, Soumali Roychowdhury, "Recognition of Channel Logos From Streamed Videos for Value Added Services in Connected TV", *IEEE International Conference for Consumer Electronics (ICCE),* Las Vegas, USA , January 2011.

[6]. T. Chattopadhyay, Arpan Pal, Utpal Garain, "Mash up of Breaking News and Contextual Web Information: A Novel Service for Connected Television", *Proceedings of 19th International Conference on Computer Communications and Networks (ICCCN),* Zurich, Switzerland, August 2010.

[7]. T. Chattopadhyay, Aniruddha Sinha, Arpan Pal, "TV Video Context Extraction", *IEEE Trends and Developments in Converging Technology towards 2020 (TENCON 2011),* Bali, INDONESIA, November 21-24, 2011.

Chapter 6

[8]. Arpan Pal, Chirabrata Bhaumik, Debnarayan Kar, Somnath Ghoshdastidar, Jasma Shukla, "A Novel On-Screen Keyboard for Hierarchical Navigation with Reduced Number of Key Strokes", *IEEE International Conference on Systems, Man and Cybernetics (SMC),* San Antonio, Texas, October 2009.

[9]. Arpan Pal, Debatri Chatterjee, Debnarayan Kar, "Evaluation and Improvements of on-screen keyboard for Television and Set-top Box", *IEEE International Symposium for Consumer Electronics (ISCE),* Singapore, June 2011.

Appendix B – List of Publications and Patents

Journals / Book Chapters

Chapter 2

[10]. Arpan Pal, M. Prashant, Avik Ghose, Chirabrata Bhaumik, "Home Infotainment Platform – A Ubiquitous Access Device for Masses", *Book Chapter in Springer Communications in Computer and Information Science,* Volume 75, 2010, Pages 11-19.DOI: 10.1007/978-3-642-13467-8.

[11]. Arpan Pal, Ramjee Prasad, Rohit Gupta, "A low-cost Connected TV platform for Emerging Markets–Requirement Analysis through User Study", *Engineering Science and Technology: An International Journal (ESTIJ),* ISSN: 2250-3498, Vol.2, No.6, December 2012.

Chapter 4

[12]. T. Chattopadhyay and Arpan Pal, "Watermarking for H.264 Video", *EE Times Design, Signal Processing Design Line,* November 2007.

Chapter 5

[13]. Arpan Pal, Aniruddha Sinha and Tanushyam Chattopadhyay, "Recognition of Characters from Streaming Videos", *Book Chapter in book: Character Recognition,* Edited by Minoru Mori, *Sciyo Publications,* ISBN: 978-953-307-105-3, September 2010.

[14]. Arpan Pal, Tanushyam Chattopadhyay, Aniruddha Sinha and Ramjee Prasad, "The Context-aware Television using Logo Detection and Character Recognition", **(Submitted)** *Springer Journal of Pattern Analysis and Applications*

Chapter 6

[15]. Debatri Chatterjee, Aniruddha Sinha, Arpan Pal, Anupam Basu,, "An Iterative Methodology to Improve TV Onscreen Keyboard Layout Design Through Evaluation of user Study*", Journal of Advances in Computing, Vol.2, No.5, October 2012),* Scientific and Academic Publishing (SAP), p-ISSN:2163-2944, e-ISSN:2163-2979.

Patents Filed

Chapter 2

[a]. Methods and apparatus for implementation of a set of Interactive Applications using a flexible Framework, 1028/MUM/2008

Chapter 3

[b]. Methods and apparatus for video conferencing solution for STB devices, 1029/MUM/2008

Chapter 4

[c]. Method And Apparatus For Watermarking, 208 /MUM/2007, **Granted in USA,** Patent no. 8,189,856, issued in May 2012

[d]. A Method And Apparatus For Encryption Of A Video Sequence, 206 /MUM/2007

Chapter 5

[e]. System for obtaining information about TV broadcasts via a plurality of channels, 2036/MUM/2009

[f]. System for SMS value added service for active television shows on a set Top Box, 2093/MUM/2009

[g]. A System and Method for Obtaining Additional Relevant Information about the Breaking News and Ticker News While Watching TV, 3039/MUM/2009

Chapter 6

[h]. Input Mechanisms, 2035/MUM/2008

Arpan Pal received both B.Tech. in Electronics and Electrical Communication Engineering and M.Tech. in Telecommunication Systems Engineering from Indian Institute of Technology, Kharagpur, India in 1990 and 1993 respectively. He has more than 20 years of experience in the area of Signal Processing., Communication and Real-time Embedded Systems.

From 1990 to 1991, he worked with APLAB Ltd. where he was involved in developing Microprocessor based Telephone Exchange Monitoring Equipment. From 1993 to 1997, he was with Research Center Imarat (RCI), a lab under Defense Research and Development Organization (DRDO) of Indian Govt. working in the area of Missile Seeker Signal Processing. From 1997 to 2002, he was leading the Real-time systems group in Rebaca Technologies (erstwhile Macmet Interactive Technologies) working in the area of Digital TV and Set top Boxes.

Since 2002, he is with Tata Consultancy Services (TCS), where he is currently heading research at Innovation Lab, Kolkata. He is also a member of Systems Research Council of TCS. His main responsibility is in conceptualizing and guiding R&D in the area of cyber-physical systems and ubiquitous computing with focus on applying the R&D outcome in the area Intelligent Infrastructure. His current research interests include Connected TV, Mobile phone and Camera based Sensing and Analytics, Physiological Sensing, M2M communications and Internet-of-Things based Applications.

He has more than 40 publications till date in reputed Journals and Conferences along with a couple of Book Chapters. He has also filed for more than 35 patents and has five patents granted to him. He serves on the Technical Program Committee of several International Conferences. He is also a reviewer for several notable journals.

Contact Address

Tata Consultancy Services Ltd. (TCS)
Bengal Intelligent Park, Bldg. - D
Plot A2, M2 & N2, Sector V, Block GP
Salt Lake Electronics Complex
Kolkata - 700091
West Bengal, India
Email: arpan.pal@tcs.com